

Глава 7

Построение доверительных интервалов

ПРИМЕНЕНИЕ СТАТИСТИКИ: аудиторская проверка накладных в компании Saxon Home Improvement Company

7.1. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПРИ ИЗВЕСТНОМ СТАНДАРТНОМ ОТКЛОНЕНИИ

Процедуры Excel: построение доверительного интервала для математического ожидания генеральной совокупности при известном стандартном отклонении

7.2. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПРИ НЕИЗВЕСТНОЙ ДИСПЕРСИИ

Распределение Стьюдента
Степени свободы
Доверительный интервал
Процедуры Excel: построение доверительного интервала для математического ожидания генеральной совокупности при неизвестном стандартном отклонении

7.3. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ ДОЛИ ПРИЗНАКА В ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Процедуры Excel: построение доверительного интервала для доли признака в генеральной совокупности

7.4. ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ

Определение объема выборки для оценки математического ожидания
Процедуры Excel: определение объема выборки для оценки математического ожидания

Определение объема выборки для оценки доли признака в генеральной совокупности
Процедуры Excel: определение объема выборки для оценки доли признака в генеральной совокупности

7.5. ПРИМЕНЕНИЕ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ В АУДИТОРСКОМ ДЕЛЕ

Оценка общей суммы элементов генеральной совокупности
Процедуры Excel: построение доверительного интервала для общей суммы генеральной совокупности
Оценка разности
Процедуры Excel: построение доверительного интервала для полной разности и генеральной совокупности
Односторонняя оценка доли нарушений установленных правил

7.6. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ И ЭТИЧЕСКИЕ ПРОБЛЕМЫ

7.7. ВЫЧИСЛЕНИЕ ОЦЕНОК И ОБЪЕМА ВЫБОРОК, ИЗВЛЕЧЕННЫХ ИЗ КОНЕЧНОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Оценка математического ожидания
Оценка доли признака
Определение объема выборки
Процедуры Excel: вычисление поправочного коэффициента для конечной генеральной совокупности

СПРАВОЧНИК ПО EXCEL. ГЛАВА 7

ЧЕМУ ДОЛЖЕН НАУЧИТЬСЯ СТУДЕНТ

- Строить доверительные интервалы для математического ожидания и доли признака.
- Определять объем выборки, необходимый для построения доверительного интервала, содержащего математическое ожидание или долю признака.
- Применять доверительные интервалы в аудиторском деле.

ПРИМЕНЕНИЕ СТАТИСТИКИ

Аудиторская проверка накладных в компании Saxon Home Improvement Company



Компания Saxon Home Improvement Company — крупный дистрибьютор санитарно-технического оборудования в пригороде одного из городов на северо-западе США. Представьте себе, что вы работаете бухгалтером этой компании и отвечаете за точность записей в информационной системе. Разумеется, оценить точность регистрации документов можно путем тотальной проверки всех записей. Однако этот путь слишком долг и дорог. Чтобы оценить точность всей генеральной совокупности записей, достаточно извлечь из нее относительно небольшую выборку. Поэтому в конце каждого месяца аудиторы извлекают из ин-

формационной системы выборку накладных и выполняют следующие процедуры.

- Вычисляют среднюю сумму, указанную в накладных за месяц.
- Подсчитывают общую сумму, указанную в накладных за месяц.
- Вычисляют все разности между суммами, указанными в накладных, и суммами, занесенными в информационную систему.
- Определяют, как часто встречаются ошибки, нарушающие систему отчетности, принятую в компании. К этим ошибкам относятся: отгрузка партии товара без складских расписок, неправильно указанные номера накладных и ошибочная отгрузка товара.

Насколько точна информация, полученная на основе анализа выборки, и как ее использовать? Достаточно ли велик объем выборки для того, чтобы делать обоснованные выводы?

ВВЕДЕНИЕ

В теории статистического вывода для оценки характеристик генеральной совокупности используются выборочные параметры. В этой главе мы рассмотрим статистические процедуры, позволяющие оценить математическое ожидание и долю признака в генеральной совокупности.

В статистике существует два вида оценок: точечные и интервальные. **Точечная оценка** (point estimate) представляет собой отдельную выборочную статистику, которая используется для оценки параметра генеральной совокупности. Например, выборочное среднее \bar{X} — это точечная оценка математического ожидания генеральной совокупности, а выборочная дисперсия S^2 — точечная оценка дисперсии генеральной совокупности σ^2 . В разделе 6.6 показано, что выборочное среднее является *несмещенной* оценкой математического ожидания генеральной совокупности. Хотя на практике, как правило, используется только одна выборка, можно показать, что среднее всех выборочных средних равно математическому ожиданию генеральной совокупности μ .¹ При оценке

¹ Аналогично, чтобы выборочная дисперсия S^2 стала несмещенной оценкой дисперсии генеральной совокупности σ , знаменатель выборочной дисперсии следует положить равным $n-1$, а не n . Иначе говоря, дисперсия генеральной совокупности является средним значением всевозможных выборочных дисперсий.

параметров генеральной совокупности следует иметь в виду, что выборочные статистики, такие как \bar{X} , зависят от конкретных выборок. Чтобы учесть этот факт, для получения **интервальной оценки** (interval estimate) математического ожидания генеральной совокупности анализируют распределение выборочных средних (см. раздел 6.6). Построенный интервал характеризуется определенным доверительным уровнем, который представляет собой вероятность того, что истинный параметр генеральной совокупности оценен правильно. Аналогичные доверительные интервалы можно применять для оценки доли признака p и основной распределенной массы генеральной совокупности. Например, можно построить доверительный интервал, которому с вероятностью, равной 95%, принадлежит средняя сумма, указанная в накладных компании Saxon Home Improvement Company за последний месяц.

В этой главе мы покажем, как определить объем выборки, чтобы гарантировать заданные свойства доверительного интервала, а также обсудим несколько важных аудиторских процедур.

7.1. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПРИ ИЗВЕСТНОМ СТАНДАРТНОМ ОТКЛОНЕНИИ

В разделе 6.6 мы применили информацию о распределении и центральную предельную теорему, чтобы вычислить процентную долю выборочных средних, лежащих в окрестности математического ожидания генеральной совокупности на расстоянии, не превышающем заданное. Например, в примере, посвященном анализу веса коробок с кукурузными хлопьями, 95% всех выборочных средних лежат в интервале от 362,12 до 373,88%. Это утверждение получено с помощью *дедукции* (deductive reasoning), поскольку свойства выборки определялись на основе данных о генеральной совокупности. Однако теперь нам необходимо применить противоположный способ рассуждения — *индуктивный* (inductive reasoning), так как в теории статистического вывода, наоборот, свойства генеральной совокупности определяются на основе параметров отдельной выборки. На практике математическое ожидание генеральной совокупности, как правило, не известно и должно быть оценено. Вернемся к примеру с коробками. Допустим теперь, что математическое ожидание генеральной совокупности μ не известно, а истинное стандартное отклонение генеральной совокупности σ равно 15 г. Тогда в формулах для определения границ интервала $\mu \pm 1,96\sigma/\sqrt{n}$ математическое ожидание генеральной совокупности μ следует заменить выборочным средним \bar{X} . Следовательно, для оценки неизвестного математического ожидания генеральной совокупности можно применить интервал $\bar{X} \pm 1,96\sigma/\sqrt{n}$. На практике, как правило, из генеральной совокупности извлекают выборку фиксированного объема n и вычисляют выборочное среднее \bar{X} . Однако, чтобы понять смысл интервальной оценки, необходимо рассмотреть гипотетическое множество всех возможных выборок данного объема.

Допустим, например, что математическое ожидание генеральной совокупности μ равно 368 г. Предположим, что из генеральной совокупности коробок извлекается выборка, состоящая из 25 коробок, средний вес которых равен 362,3 г. Интервал, предложенный для оценки математического ожидания μ , имеет границы: $362,3 \pm 1,96 \times 15/\sqrt{25}$, т.е. $362,3 \pm 5,88$. Таким образом,

$$356,42 \leq \mu \leq 368,18.$$

Поскольку математическое ожидание генеральной совокупности μ равно 368, оно попадает в этот интервал, и, следовательно, данная выборка позволяет правильно ее оценить (рис. 7.1).

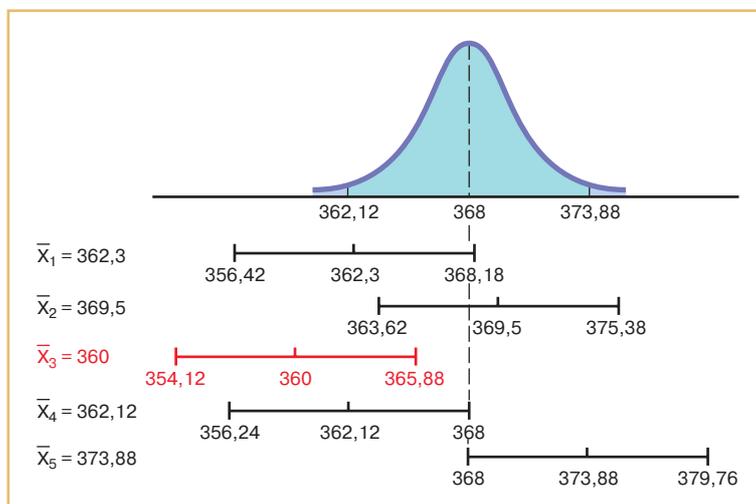


Рис. 7.1. Доверительные интервальные оценки математического ожидания генеральной совокупности, полученные по пяти разным выборкам объема $n = 25$, извлеченным из генеральной совокупности с параметрами $\mu = 368$ и $\sigma = 15$

Продолжим анализ нашего гипотетического примера. Допустим, что выборочное среднее некоей выборки объема $n = 25$, равно **369,5**. Доверительный интервал, построенный по этой выборке, имеет границы, равные $369,5 \pm 1,96 \times 15 / \sqrt{25}$, т.е. **369,5 ± 5,88**. Таким образом,

$$363,62 \leq \mu \leq 375,38.$$

Поскольку математическое ожидание генеральной совокупности μ равно 368, оно попадает в этот интервал, и, следовательно, данная оценка является правильной.

Возникает впечатление, что выборки, имеющие объем $n = 25$, всегда приводят к правильным оценкам математического ожидания генеральной совокупности μ . Чтобы опровергнуть это, рассмотрим третий гипотетический пример. Допустим, что средний вес коробок равен 360 г. Интервал, предложенный для оценки математического ожидания μ , имеет границы: $360 \pm 1,96 \times 15 / \sqrt{25}$, т.е. **360 ± 5,88**. Таким образом, в данном случае

$$354,12 \leq \mu \leq 365,88.$$

Эта оценка *неверна*, поскольку математическое ожидание генеральной совокупности μ равно 368 и не попадает в этот интервал (см. рис. 7.1). Таким образом, для некоторых выборок эта оценка верна, а для некоторых — нет. Кроме того, на практике, как правило, из генеральной совокупности извлекается только одна выборка. Следовательно, поскольку математическое ожидание генеральной совокупности μ не известно, невозможно сказать, верна полученная интервальная оценка или нет.

Чтобы разрешить эту дилемму, необходимо определить долю выборок, позволяющих правильно оценить математическое ожидание генеральной совокупности μ . Для этого следует исследовать еще две гипотетические выборки, средние значения которых равны 362,12 и 373,88 г соответственно.

Если $\bar{X} = 362,12$, мы получаем интервал $362,12 \pm 1,96 \times 15 / \sqrt{25}$, т.е. **362,12 ± 5,88**. Это приводит к оценке

$$356,24 \leq \mu \leq 368,00.$$

Поскольку математическое ожидание генеральной совокупности μ , равное 368, является верхней границей интервала, эта оценка верна.

Если $\bar{X} = 373,88$, мы получаем интервал $373,88 \pm 1,96 \times 15 / \sqrt{25}$, т.е. $373,88 \pm 5,88$. Это приводит к оценке

$$368,00 \leq \mu \leq 379,76.$$

Поскольку математическое ожидание генеральной совокупности μ , равное 368, является нижней границей интервала, эта оценка верна.

Таким образом, если выборочное среднее изменяется в диапазоне от 362,12 до 373,88 г, математическое ожидание генеральной совокупности лежит *где-то* внутри этого соответствующего доверительного интервала. Вероятность того, что это значение лежит в интервале с границами 362,12 и 373,88, равна 95% (см. раздел 6.6). Следовательно, 95% средних значений всех выборок, имеющих объем $n = 25$, позволяют правильно оценить математическое ожидание генеральной совокупности, а 5% — нет.

Интервальная оценка, доверительный уровень которой равен 95%, интерпретируется следующим образом: если из генеральной совокупности извлечь все выборки, имеющие объем n , и вычислить их выборочные средние, то 95% доверительных интервалов, построенных на их основе, будут содержать математическое ожидание генеральной совокупности, а 5% — нет.

На практике, как правило, из генеральной совокупности извлекается только одна выборка, а математическое ожидание генеральной совокупности μ не известно. По этой причине невозможно гарантировать, что некий конкретный доверительный интервал содержит величину μ . Можно лишь утверждать, что вероятность этого события равна 95%.

В некоторых ситуациях желательно иметь более высокий доверительный уровень, а следовательно, точность оценки величины μ (например, 99%). Но иногда можно ограничиться и менее точной оценкой (например, 90%).

Как правило, **доверительный уровень** (level of confidence) обозначают следующим образом: $(1-\alpha) \times 100\%$, где величина α представляет собой площадь, ограниченную хвостом распределения, выходящим за пределы доверительного интервала. (Величину α называют *уровнем значимости* доверительного интервала. Кроме того, в качестве синонима для доверительного уровня иногда употребляется выражение “доверительная вероятность”. — Прим. ред.) Площади, ограниченные как левым, так и правым хвостами распределения, выходящими за пределы доверительного интервала, равны $\alpha/2$.

**ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ МАТЕМАТИЧЕСКОГО
ОЖИДАНИЯ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПРИ ИЗВЕСТНОМ
СТАНДАРТНОМ ОТКЛОНЕНИИ**

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}},$$

или

$$\bar{X} - Z \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z \frac{\sigma}{\sqrt{n}}, \quad (7.1)$$

где Z — значение стандартизованной нормально распределенной случайной величины, соответствующее интегральной вероятности, равной $1 - \alpha/2$, σ — стандартное отклонение генеральной совокупности.

Величина Z , выбранная для построения доверительного интервала, называется **критическим значением** (critical value) распределения. Чтобы построить интервал, имеющий 95%-ный доверительный уровень, необходимо выбрать $\alpha = 0,05$. Половина площади симметричной фигуры, ограниченной концами доверительного интервала и кривой стандартизованного нормального распределения, равна $0,95/2=0,4750$. Величина Z , соответствующая площади 0,4750, равна 1,96, поскольку площадь фигуры, ограниченной правым хвостом распределения, равна 0,025, а суммарная площадь фигуры, лежащей левее значения $Z = 1,96$, равна 0,975.

Каждому доверительному уровню $1 - \alpha$ соответствует свое критическое значение. Например, доверительному уровню, равному 95%, соответствует $Z = \pm 1,96$ (рис. 7.2). Если требуется построить интервал, доверительный уровень которого равен 99%, следует выбрать $\alpha = 0,01$. В этом случае величина Z приблизительно равна 2,58, поскольку площадь, ограниченная правым хвостом распределения, выходящим за пределы доверительного интервала, равна 0,005, а суммарная площадь фигуры, лежащей левее значения Z , равна 0,995 (рис. 7.3).

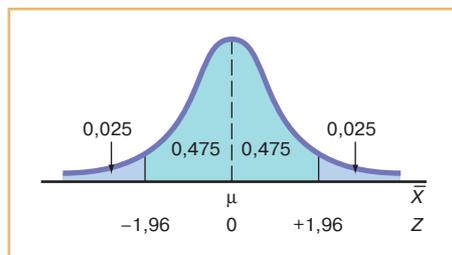


Рис. 7.2. Гауссова кривая для определения критического значения Z , соответствующего доверительному уровню, равному 95%

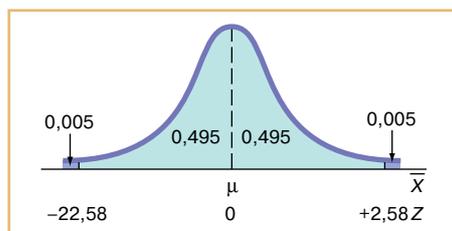


Рис. 7.3. Гауссова кривая для определения критического значения Z , соответствующего доверительному уровню, равному 99%

Возникает вопрос, почему бы не построить интервал, доверительный уровень которого был бы очень близок к 100%. Это нецелесообразно, поскольку такой доверительный интервал оказался бы слишком широким, а оценка математического ожидания — слишком неточной. Разумеется, вероятность того, что математическое ожидание лежит в этом интервале, очень высока, однако для принятия решения этот факт практически бесполезен. В разделе 7.4 мы более подробно обсудим способы достижения компромисса между шириной интервала и доверительным уровнем.

ПРИМЕР 7.1. ОЦЕНКА СРЕДНЕЙ ДЛИНЫ ЛИСТА БУМАГИ С ДОВЕРИТЕЛЬНЫМ УРОВНЕМ, РАВНЫМ 95%

При производстве бумаги средняя длина листа должна быть равной 11 дюймам, а ее стандартное отклонение — 0,02 дюйма. Периодически из произведенной продукции, чтобы оценить ее качество, извлекаются выборки. Допустим, выборка состоит из 100 листов, а ее выборочное среднее — 10,998 дюйма. Постройте интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 95%.

РЕШЕНИЕ. Подставим в формулу (7.1) величину $Z = 1,96$, соответствующую доверительному уровню, равному 95%.

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}} \leq 10,998 \pm 1,96 \frac{0,02}{\sqrt{100}} = 10,998 \pm 0,00392 .$$

$$10,99408 \leq \mu \leq 11,00192 .$$

Таким образом, вероятность того, что математическое ожидание генеральной совокупности лежит в интервале от 10,99408 до 11,00192, равна 95%. Поскольку номинальная длина бумаги — 11 дюймов, она попадает в построенный интервал. Следовательно, производственный процесс выполняется правильно. ■

ПРИМЕР 7.2. ОЦЕНКА СРЕДНЕЙ ДЛИНЫ ЛИСТА БУМАГИ С ДОВЕРИТЕЛЬНЫМ УРОВНЕМ, РАВНЫМ 99%

Постройте интервал, содержащий математическое ожидание генеральной совокупности листов бумаги, доверительный уровень которого равен 99%.

РЕШЕНИЕ. Подставим в формулу (7.1) величину $Z = 2,58$, соответствующую доверительному уровню, равному 99%.

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}} \leq 10,998 \pm 2,58 \frac{0,02}{\sqrt{100}} = 10,998 \pm 0,00516 .$$

$$10,99284 \leq \mu \leq 11,00316 .$$

Поскольку номинальная длина бумаги попадает в построенный интервал, можно сделать вывод, что производственный процесс выполняется правильно. ■

Процедуры Excel: построение доверительного интервала для математического ожидания генеральной совокупности при известном стандартном отклонении

Чтобы построить доверительный интервал для математического ожидания генеральной совокупности по известному стандартному отклонению, можно воспользоваться функцией ДОВЕРИТ или применить надстройку PHStat2.

Например, чтобы построить доверительный интервал для средней длины листа бумаги в примере 7.1, необходимо воспользоваться одной из следующих процедур.

Применение Excel в сочетании с надстройкой PHStat2

Чтобы построить доверительный интервал для математического ожидания генеральной совокупности по известному стандартному отклонению, следует применить процедуру PHStat⇒Confidence Intervals⇒Estimate for the Mean, sigma known... (PHStat⇒Доверительные интервалы⇒Оценка математического ожидания, стандартное отклонение известно...).

1. Выбрать команду PHStat⇒Confidence Intervals⇒Estimate for the Mean, sigma known.
2. В диалоговом окне Estimate for the Mean, sigma known (см. иллюстрацию) сделать следующее.
 - 2.1. Ввести в окне редактирования Population Standard Deviation (Стандартное отклонение генеральной совокупности) число 0.02.
 - 2.2. Ввести в окне редактирования Confidence Level (Доверительный уровень) число 95.
 - 2.3. Установить переключатель Input Options (Параметры ввода) в положение Sample Statistics Known (Выборочные статистики известны) и ввести в окне редактирования Sample Size (Объем выборки) число 100, а в окне редактирования Sample Mean (Выборочное среднее) – число 10.998.
 - 2.4. Ввести в окне редактирования Title соответствующий заголовок.
 - 2.5. Щелкнуть на кнопке ОК.



Если выборочное среднее неизвестно, в п. 2.3 следует установить переключатель Input Options в положение Sample Statistics Unknown (Выборочные статистики неизвестны), а в окне редактирования Sample Cell Range (Диапазон ячеек, содержащий выборку) ввести диапазон ячеек, в которых записаны элементы выборки.

Применение Excel

Инструкции, позволяющие самостоятельно создать рабочий лист, вычисляющий доверительный интервал для математического ожидания генеральной совокупности по известному стандартному отклонению, содержатся в разделе ЕН.7.1.

 Chapter 7.xls

Данные, на основе которых вычисляется доверительный интервал, содержащий среднюю длину листа бумаги по известному стандартному отклонению (пример 7.1), хранятся в рабочей книге Chapter 7.xls на листе Пример7.1.

УПРАЖНЕНИЯ К РАЗДЕЛУ 7.1**Изучение основ**

- 7.1. Предположим, что $\bar{X} = 85$, $\sigma = 8$ и $n = 64$. Постройте 95%-ный доверительный интервал для математического ожидания генеральной совокупности.
- 7.2. Предположим, что $\bar{X} = 125$, $\sigma = 24$ и $n = 36$. Постройте 99%-ный доверительный интервал для математического ожидания генеральной совокупности.
- 7.3. По данным специалиста по маркетингу, вероятность того, что средний объем продаж колеблется между 170 000 и 200 000 долл., равна 95%. Объясните смысл этого утверждения.
- 7.4. Почему в примере 7.1 невозможно добиться 100%-ного доверительного уровня?
- 7.5. Правда ли, что 95% всех выборочных средних, вычисленных в примере 7.1, лежат в интервале от 10,99408 до 11,00192 дюймов? Обоснуйте свой ответ.

- 7.6. Правда ли, что существуют выборочные средние, вычисленные в примере 7.1, не лежащие между числами 10,99408 и 11,00192 дюймов? Обоснуйте свой ответ.

Применение понятий

- 7.7. Менеджер магазина стройматериалов хочет оценить фактический объем краски, содержащейся в галлонных банках известной компании. Известно, что стандартное отклонение объема краски равно 0,02 галлона. Менеджер выбрал 50 банок. Выборочный средний объем равен 0,995 галлона.
1. Постройте интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 99%.
 2. Основываясь на решении задачи 1, определите, стоит ли менеджеру жаловаться на завод-производитель и почему.
 3. Можно ли утверждать, что объем краски в банке подчиняется нормальному закону распределения? Объясните свой ответ.
 4. Объясните, почему объем краски, равный 0,98 галлона, считается допустимым, хотя он не попадает в построенный вами доверительный интервал.
 5. Допустим, что доверительный уровень интервала должен быть равен 95%. Как изменятся ответы на вопросы 1 и 2?
- 7.8. Менеджер из отдела контроля за качеством продукции на заводе, производящем электрические лампочки, желает оценить среднюю продолжительность работы лампочек из крупной партии. Номинальное стандартное отклонение равно 100 ч. Для контроля выбрана партия, состоящая из 64 лампочек, средняя продолжительность работы которых равна 350 ч.
1. Постройте интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 95%.
 2. Можно ли утверждать, что средняя продолжительность работы лампочек из анализируемой партии не меньше 400 ч? Обоснуйте свой ответ.
 3. Можно ли утверждать, что продолжительность работы лампочек подчиняется нормальному закону распределения? Объясните свой ответ.
 4. Объясните, почему продолжительность работы лампочки, равная 320 ч, считается допустимой, хотя она не попадает в построенный вами доверительный интервал.
 5. Допустим, что стандартное отклонение продолжительности работы лампочек равно 80 ч. Как изменятся ответы на вопросы 1 и 2?
- 7.9. Отдел технического контроля на заводе, производящем газированные напитки, желает выяснить фактический объем жидкости, содержащейся в двухлитровых бутылках. Известно, что стандартное отклонение объема жидкости в двухлитровой бутылке равно 0,05 л. Менеджер выбрал 100 двухлитровых бутылок. Выборочный средний объем составил 1,99 л.
1. Постройте интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 95%.
 2. Можно ли утверждать, что объем жидкости в бутылках подчиняется нормальному закону распределения? Объясните свой ответ.
 3. Объясните, почему объем жидкости, равный 2,02 л, считается допустимым, хотя он не попадает в построенный вами доверительный интервал.
 4. Предположим, что выборочное среднее равно 1,97 л. Как изменится ответ на вопрос 1?
-

7.2. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПРИ НЕИЗВЕСТНОЙ ДИСПЕРСИИ

На практике как математическое значение генеральной совокупности, так и его стандартное отклонение часто бывают неизвестными. Следовательно, необходимо построить доверительный интервал, содержащий математическое значение генеральной совокупности, используя лишь выборочные статистики \bar{X} и S .

Распределение Стьюдента

В начале 20-го века статистик Уильям С. Госсет (William S. Gosset), сотрудник ирландского отделения пивоваренной компании Guinness [3], заинтересовался проблемой оценки математического ожидания при неизвестном стандартном отклонении. Поскольку компания Guinness запрещала своим сотрудникам публиковать работы под собственными именами, Госсет взял псевдоним *Стьюдент*. По этой причине распределение, предложенное Госсетом, называется *t-распределением Стьюдента* (Student's *t* distribution).

Если случайная величина X является нормально распределенной, то следующая статистика имеет *t*-распределение с $n-1$ степенями свободы:

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}.$$

Обратите внимание на то, что это выражение почти совпадает с формулой (6.12), только вместо величины σ в нем стоит выборочное стандартное отклонение S . Понятие степени свободы обсуждается далее.

Внешне распределение Стьюдента очень напоминает стандартизованное нормальное распределение. Оба распределения имеют колоколообразную форму и являются симметричными. Однако хвосты *t*-распределения “тяжелее” (т.е. ограничивают большую площадь), а площадь фигуры в центре определения меньше, чем у стандартизованного нормального распределения (рис. 7.4). Это происходит потому, что стандартное отклонение σ не известно, а вместо него используется его выборочная оценка S . Неопределенность значения σ порождает большую изменчивость переменной t по сравнению с величиной Z .

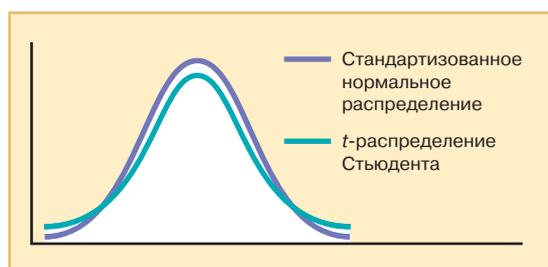


Рис. 7.4. Стандартизованное нормальное распределение и *t*-распределение Стьюдента с пятью степенями свободы

Однако при увеличении количества степеней свободы *t*-распределение становится все ближе к стандартизованному нормальному распределению. Это происходит потому, что при увеличении объема выборки оценка S становится все точнее. При объеме выборки, равном 120 и более, величина S довольно точно аппроксимирует стандартное отклонение σ , так что разница между *t*-распределением и стандартизованным нормальным отклонением становится минимальной. По этой причине, если объем выборки превышает 120, многие статистики вместо величины t используют переменную Z .

ПРОВЕРКА ПРЕДПОЛОЖЕНИЙ

Напомним, что t -распределение основано на предположении, что изучаемая случайная величина X является нормально распределенной. Однако на практике t -распределение можно применять для оценки неизвестного математического ожидания генеральной совокупности при неизвестном стандартном отклонении при достаточно большом объеме выборки и не слишком асимметричном распределении. При работе с небольшими выборками эти условия уже не выполняются автоматически, поэтому их следует проверять. Для этого необходимо строить гистограмму, диаграмму “ствол и листья”, блочную диаграмму или график нормального распределения.

Критические значения для t -распределения с соответствующими степенями свободы можно найти в табл. Д.3. В заголовке каждого столбца этой таблицы указана площадь фигуры, ограниченной хвостом t -распределения (поскольку для переменной t указаны только положительные значения, эта площадь ограничена правым хвостом распределения). В заголовке каждой строки указано количество степеней свободы. Например, в табл. 7.1 показано, как найти площадь фигуры, ограниченной t -распределением, имеющим 99 степеней свободы, и соответствующим значением переменной t , если необходимо построить интервал, доверительный уровень которого равен 95%. Этот доверительный уровень означает, что площадь каждой фигуры, ограниченной хвостами t -распределения, равна 0,025. Найдем пересечение столбца, соответствующего величине 0,025, и строки, соответствующей 99 степеням свободы. В этой ячейке записано критическое значение, равное 1,9842. Поскольку t -распределение является симметричным и его математическое ожидание равно 0, площади, ограниченной правым хвостом, соответствует величина +1,9842, а площади, ограниченной левым хвостом, соответствует величина -1,9842. Величина 1,9842 означает следующее: вероятность того, что величина t превосходит +1,9842, равна 0,025, т.е. 2,5% (рис. 7.5).

Таблица 7.1.² Определение площади фигуры, ограниченной t -распределением, имеющим 99 степеней свободы, и соответствующим значением переменной t

Площади фигуры, ограниченной правым хвостом t -распределения						
Степени свободы	0,25	0,10	0,05	0,025	0,01	0,005
1	1,0000	3,0777	6,3138	12,7062	31,8207	63,6574
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,7649	1,6477	2,3534	3,1824	4,5407	5,8409
4	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0322
...
...
...
96	0,6771	1,2904	1,6609	1,9850	2,3658	2,6280
97	0,6770	1,2903	1,6607	1,9847	2,3654	2,6275
98	0,6770	1,2902	1,6606	1,9845	2,3650	2,6269
99	0,6770	1,2902	1,6604	1,9842	2,3646	2,6264
100	0,6770	1,2901	1,6602	1,9840	2,3642	2,6259

² Фрагмент табл. Д.3.

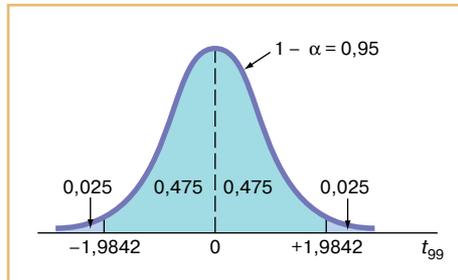


Рис. 7.5. Распределение Стьюдента с 99 степенями свободы

Степени свободы

Напомним, что для вычисления выборочной дисперсии S^2 необходимо вычислить величину

$$\sum_{i=1}^n (X_i - \bar{X})^2.$$

Таким образом, для вычисления выборочной дисперсии необходимо знать \bar{X} . Следовательно, мы можем варьировать лишь $n-1$ выборочных значений. Это означает, что величина обладает $n-1$ степенями свободы.

Допустим, например, что выборка состоит из 5 чисел, а ее выборочное среднее равно 20. Сколько разных значений необходимо знать для того, чтобы однозначно определить остальные? Если $n = 5$ и $\bar{X} = 20$, то

$$\sum_{i=1}^n X_i = 100,$$

поскольку

$$\frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Таким образом, если известны четыре величины, пятое значение уже *не* свободно, поскольку сумма должна быть равна 100. Например, если нам известны величины 18, 24, 19 и 16, пятая величина должна быть равной 23, поскольку сумма равна 100.

Доверительный интервал

Рассмотрим формулу для вычисления интервала, содержащего математическое ожидание при неизвестном стандартном отклонении с вероятностью $(1-\alpha) \times 100\%$.

ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПРИ НЕИЗВЕСТНОМ СТАНДАРТНОМ ОТКЛОНЕНИИ

$$\bar{X} + t_{n-1} \frac{\sigma}{\sqrt{n}}$$

или

$$\bar{X} - t_{n-1} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1} \frac{\sigma}{\sqrt{n}}, \quad (7.2)$$

где t_{n-1} — критическое значение t -распределения с $n-1$ степенями свободы, соответствующее площади, ограниченной правым хвостом и равной $\alpha/2$.

7.2. Построение доверительного интервала для математического ожидания... 459

Чтобы проиллюстрировать применение этой формулы, вернемся к задаче об аудиторской проверке в компании Saxon Home Improvement Company. Предположим, что из информационной системы извлечена выборка, состоящая из 100 накладных, заполненных в течение последнего месяца. Допустим, что выборочное среднее равно 110,27 долл., а выборочное стандартное отклонение — 28,95 долл. Если компания желает построить интервал, имеющий доверительный уровень, равный 95%, критическое значение t -распределения равно 1,9842 (см. табл. 7.1). Используя формулу (7.2), получаем:

$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}} = 110,27 \pm 1,9842 \frac{28,95}{\sqrt{100}} = 110,27 \pm 5,74 ,$$

$$104,53 \leq \mu \leq 116,01.$$

На рис. 7.6 продемонстрировано вычисление доверительного интервала, содержащего среднюю сумму накладных, с помощью программы Microsoft Excel.

	A	B
1	Средняя сумма накладных	
2		
3	Data	
4	Sample Standard Deviation	28.95
5	Sample Mean	110.27
6	Sample Size	100
7	Confidence Level	95%
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	2.895
11	Degrees of Freedom	99
12	t Value	1.984217305
13	Interval Half Width	5.74309101
14		
15	Confidence Interval	
16	Interval Lower Limit	104.53
17	Interval Upper Limit	116.01

Рис. 7.6. Вычисление доверительного интервала, содержащего среднюю сумму накладных, с помощью программы Microsoft Excel

Таким образом, вероятность того, что средняя сумма накладных находится в интервале от 104,53 до 116,01, равна 95%. Это значит, что если мы извлечем из информационной системы все возможные выборки, состоящие из 100 накладных (что практически невозможно), 95% доверительных интервалов будут содержать математическое ожидание генеральной совокупности. Корректность этих доверительных интервалов зависит от того, насколько распределение генеральной совокупности близко к нормальному. Поскольку объем выборки довольно велик ($n = 100$), предположение о нормальном распределении вполне правдоподобно, а полученная оценка математического ожидания довольно надежна. Таким образом, применение t -распределения вполне оправданно. На практике корректность доверительного интервала необходимо проверить, сравнив со средней суммой накладных, введенных в информационную систему. Если обнаружится расхождение, следует продолжить исследование.

Процедуры Excel: построение доверительного интервала для математического ожидания генеральной совокупности при неизвестном стандартном отклонении

Чтобы построить доверительный интервал для математического ожидания генеральной совокупности при неизвестном стандартном отклонении, можно воспользоваться функцией СТЬЮДРАСПОБЕ или надстройкой PHStat2.

Например, чтобы построить доверительный интервал для средней суммы накладных в компании Saxon Home Improvement Company, необходимо выполнить одну из следующих процедур.

Применение Excel в сочетании с надстройкой PHStat2

Чтобы построить доверительный интервал для математического ожидания генеральной совокупности при неизвестном стандартном отклонении, следует применить процедуру PHStat⇒Confidence Intervals⇒Estimate for the Mean, sigma unknown... (PHStat⇒Доверительные интервалы⇒Оценка математического ожидания, стандартное отклонение неизвестно...).

1. Выбрать PHStat⇒Confidence Intervals⇒Estimate for the Mean, sigma unknown...
2. В диалоговом окне Estimate for the Mean, sigma unknown (см. иллюстрацию) сделать следующее.
 - 2.1. Ввести в окне редактирования Confidence Level (Доверительный уровень) число 95.
 - 2.2. Установить переключатель Input Options (Параметры ввода) в положение Sample Statistics Known (Выборочные статистики известны) и ввести в окне редактирования Sample Size (Объем выборки) число 100, в окне редактирования Sample Mean (Выборочное среднее) – число 110.27, а в окне редактирования Sample Std. Deviation – число 28.95.
 - 2.3. Ввести в окне редактирования Title соответствующий заголовок.
 - 2.4. Щелкнуть на кнопке ОК.



Если выборочное среднее и выборочное стандартное отклонение не известны, в п. 2.2 следует установить переключатель Input Options в положение Sample Statistics Unknown (Выборочные статистики неизвестны), а в окне редактирования Sample Cell Range (Диапазон ячеек, содержащий выборку) ввести диапазон ячеек, в которых записаны элементы выборки.

Применение Excel

Для того чтобы самостоятельно создать рабочий лист, вычисляющий доверительный интервал для математического ожидания генеральной совокупности при неизвестном стандартном отклонении, следуйте инструкциям, приведенным в разделе ЕН.7.2.



Chapter 7.xls

Данные, на основе которых вычисляется доверительный интервал, содержащий среднюю сумму накладных, хранятся в рабочей книге Chapter 7.xls на листе Рис7.6.

ПРИМЕР 7.3. ОЦЕНКА СРЕДНЕЙ ВЕЛИЧИНЫ СИЛЫ РАЗРУШЕНИЯ ИЗОЛЯТОРА

Некая промышленная компания на Среднем Западе производит электрические изоляторы. Если во время работы изолятор выходит из строя, происходит короткое замыкание. Чтобы проверить прочность изолятора, компания проводит испытания, в ходе которых определяется максимальная сила, необходимая для разрушения изолятора. Сила измеряется в фунтах нагрузки, приводящей к разрушению изолятора. В табл. 7.2 приведены результаты 30 экспериментов.

Таблица 7.2. Сила (в фунтах), необходимая для разрушения изолятора

1 870	1 728	1 656	1 610	1 634	1 784	1 522	1 696	1 592	1 662
1 866	1 764	1 734	1 662	1 734	1 774	1 550	1 756	1 762	1 866
1 820	1 744	1 788	1 688	1 810	1 752	1 680	1 810	1 652	1 736

Постройте интервал, содержащий математическое ожидание генеральной совокупности величин силы, необходимой для разрушения изолятора, доверительный уровень которого равен 95%.

РЕШЕНИЕ. Как показано на рис. 7.7, выборочное среднее равно $\bar{X} = 1\,723,4$ фунта, а выборочное стандартное отклонение равно $S = 89,55$ фунта.

	A	B
1	Средняя прочность изоляторов	
2		
3	Data	
4	Sample Standard Deviation	89.55083319
5	Sample Mean	1723.4
6	Sample Size	30
7	Confidence Level	95%
8	Intermediate Calculations	
9	Standard Error of the Mean	16.3897046
10	Degrees of Freedom	29
11	t Value	2.045230756
12	Interval Half Width	33.4384891
13	Confidence Interval	
14	Interval Lower Limit	1689.96
15	Interval Upper Limit	1756.84

Рис. 7.7. Вычисление доверительного интервала для средней величины силы, необходимой для разрушения изолятора, с помощью программы Microsoft Excel

Чтобы вычислить нижнюю и верхнюю границы доверительного интервала с помощью формулы (7.2), необходимо определить критическое значение t -распределения с 29 степенями свободы для площади, равной 0,025. Как следует из табл. Д.3, критическое значение $t_{29} = 2,0452$.

Итак, $\bar{X} = 1\,723,4$, $S = 89,55$, $n = 30$ и $t_{29} = 2,0452$. Следовательно,

$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}} = 1\,723,4 \pm 2,0452 \frac{89,55}{\sqrt{30}} = 1\,723,4 \pm 33,44.$$

$$1\,689,96 \leq \mu \leq 1\,756,84.$$

Таким образом, вероятность того, что средняя величина силы разрушения изолятора находится в интервале от 1 689,96 до 1 756,84, равна 95%. Корректность этих доверительных интервалов зависит от того, насколько распределение генеральной совокупности близко к нормальному. Напомним, что требование о большом объеме выборки можно ослабить. Таким образом, если объем выборки равен 30, предположение о нормальном распределении остается правдоподобным, даже если распределение силы разрушения слегка асимметрично. График нормального распределения, изображенный на рис. 7.8, и блочная диаграмма, представленная на рис. 7.9, свидетельствуют о том, что распределение силы разрушения имеет легкую асимметрию, следовательно, для решения задачи можно применять t -распределение. ■

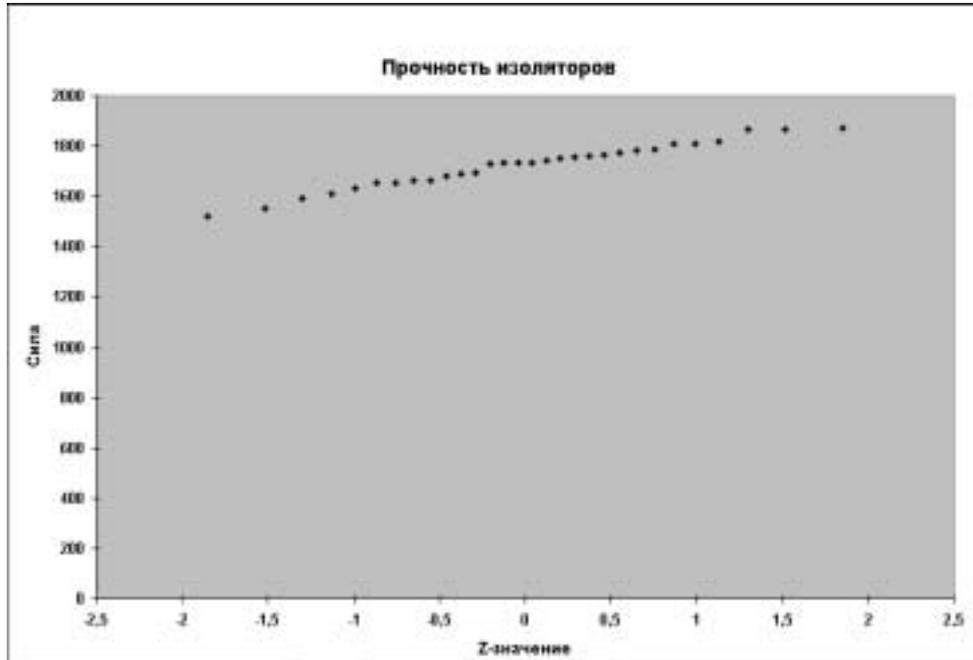


Рис. 7.8. График нормального распределения для средней величины силы, необходимой для разрушения изолятора, построенный с помощью программы Microsoft Excel

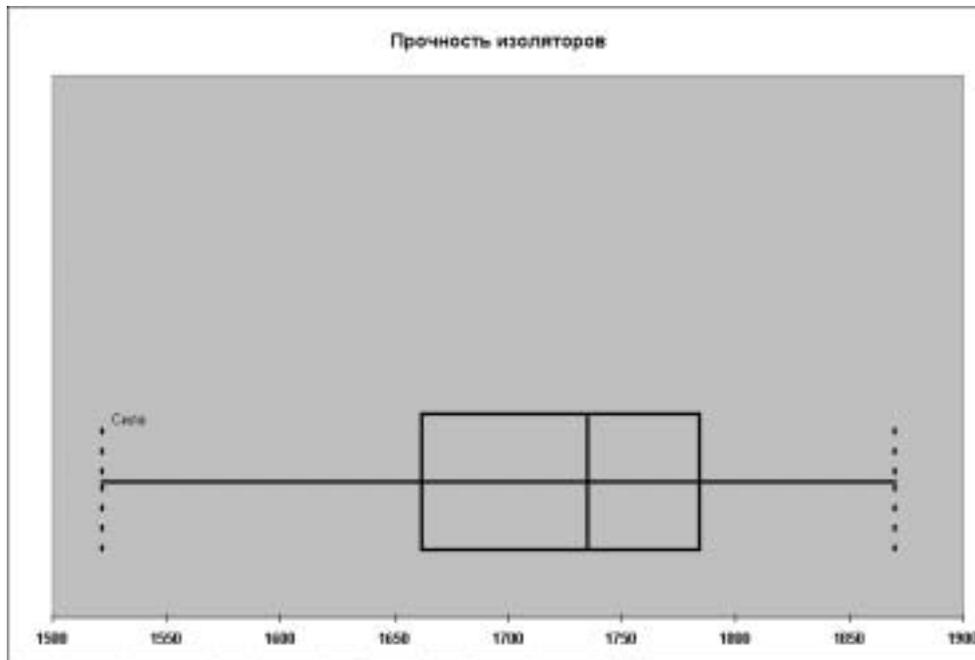


Рис. 7.9. Блочная диаграмма для средней величины силы, необходимой для разрушения изолятора, построенная с помощью программы Microsoft Excel

УПРАЖНЕНИЯ К РАЗДЕЛУ 7.2

Изучение основ

7.10. Определите критическое значение t при следующих данных.

1. $1-\alpha = 0,95, n = 10$.
2. $1-\alpha = 0,99, n = 10$.
3. $1-\alpha = 0,95, n = 32$.
4. $1-\alpha = 0,95, n = 65$.
5. $1-\alpha = 0,90, n = 16$.

7.11. Предположим, что $\bar{X} = 75, S = 24, n = 36$, и генеральная совокупность является нормально распределенной. Постройте интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 95%.

7.12. Предположим, что $\bar{X} = 50, S = 15, n = 16$, и генеральная совокупность является нормально распределенной. Постройте интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 99%.

7.13. По каждой из выборок, приведенных ниже, постройте интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 95%, предполагая, что генеральная совокупность имеет нормальное распределение.

Выборка 1: 1 1 1 1 8 8 8 8

Выборка 2: 1 2 3 4 5 6 7 8

Объясните, почему эти выборки имеют разный доверительный интервал, хотя их средние значения и размах совпадают.

7.14. Постройте интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 95%, для выборки {1, 2, 3, 4, 5, 6, 20}. Замените 20 на 7 и снова постройте доверительный интервал. Используя этот пример, продемонстрируйте влияние выброса (т.е. экстремального значения) на доверительный интервал.

Применение понятий

Задачи 7.15–7.19 можно решать как вручную, так и с помощью программы Microsoft Excel. Задачи 7.20 и 7.21 рекомендуется решать с помощью программы Microsoft Excel.

7.15. Министерство транспорта США потребовало от производителей автомобильных шин опубликовать данные о боковинах шин, чтобы информировать покупателей о качестве товара. Одним из основных показателей качества шины является износостойчивость протектора. Этот показатель является относительным. В качестве эталона выбирается шина, у которой показатель износостойчивости равен числу 100. Таким образом, шина с показателем износостойчивости, равным 200, в среднем прослужит вдвое больше, чем эталон. Общество потребителей желает оценить реальный показатель износостойчивости шин, произведенных некоей широко известной компанией. Производитель утверждает, что показатель износостойчивости его шин равен 200. Выборочное среднее, вычисленное по случайной выборке, состоящей из 18 шин, равно 195,3, а выборочное стандартное отклонение — 21,4.

1. Предположим, что генеральная совокупность показателей износоустойчивости шин имеет нормальное распределение. Постройте интервал, содержащий математическое ожидание этой генеральной совокупности, доверительный уровень которого равен 95%.
 2. Стоит ли обществу потребителей обвинять компанию в сокрытии информации об износоустойчивости ее шин? Обоснуйте свой ответ.
 3. Объясните, почему показатель износоустойчивости шины, равный 210, не является экстремальным значением, хотя он и не попадает в доверительный интервал, построенный при решении задачи 1.
- 7.16.** Магазин канцтоваров желает оценить среднюю стоимость поздравительных открыток, хранящихся на складе. Выборочное среднее, вычисленное по случайной выборке, состоящей из 20 поздравительных открыток, равно 1,67 долл., а выборочное стандартное отклонение — 0,32 долл.
1. Предположим, что генеральная совокупность цен имеет нормальное распределение. Постройте интервал, содержащий математическое ожидание этой генеральной совокупности, доверительный уровень которого равен 95%.
 2. Может ли ответ на вопрос 1 помочь владельцу магазина оценить среднюю стоимость товаров, хранящихся на его складе?
- 7.17.** Один из основных критериев качества услуг, предоставляемых любой организацией, — скорость, с которой она реагирует на жалобы клиентов. Крупный универсам, торгующий фурнитурой и коврами, за последние годы значительно расширился. В частности, отдел ковровых покрытий, в котором прежде работали 2 человека, теперь состоит из руководителя, измерителя и 15 продавцов. На протяжении последнего года компания получила 50 жалоб на работу этого отдела. Ниже приведены данные о количестве дней, прошедших со дня получения жалобы до принятия решения.  FURNITURE.XLS.

54	5	35	137	31	27	152	2	123	81	74	27
11	19	126	110	110	29	61	35	94	31	26	5
12	4	165	32	29	28	29	26	25	1	14	13
13	10	5	27	4	52	30	22	36	26	20	23
33	68										

1. Постройте интервал, содержащий математическое ожидание этой генеральной совокупности, доверительный уровень которого равен 95%.
 2. Какие предположения должны выполняться при решении задачи 1?
 3. Можно ли утверждать, что предположения, сформулированные в задаче 2, серьезно нарушаются? Аргументируйте свой ответ.
 4. Как ответ на вопрос 3 влияет на оценку корректности доверительного интервала, построенного в задаче 1?
- 7.18.** В штате Нью-Йорк сберегательным банкам разрешено осуществлять страховые жизни. В процедуру оформления страховки входят изучение запроса, проверка медицинской информации, дополнительные медицинские исследования и проверка информации, поступившей из полиции. Чтобы страхование жизни было прибыльным для банка, необходимо ускорить оформление страховки. Банк создал выборку, в которой указано время, затраченное на оформление 27 страховок в течение одного месяца.  INSURANCE.XLS.

73	19	16	64	28	28	31	90	60	56	31	56	22	18
45	48	17	17	17	91	92	63	50	51	69	16	17	

7.2. Построение доверительного интервала для математического ожидания... 465

1. Постройте 95%-ный доверительный интервал, содержащий среднее время оформления страховки.
2. Какие предположения должны выполняться при решении задачи 1?
3. Можно ли утверждать, что предположения, сформулированные в задаче 2, серьезно нарушаются? Аргументируйте свой ответ.
4. Сравните выводы, полученные при решении задачи 1, с результатами решения задачи 3.56.

7.19. Стоимость номера в гостинице и проката автомобилей в разных городах различается. В таблице представлены данные о 20 городах, опубликованные компанией Dow Jones Travel Index по состоянию на 2–4 июня 2002 года.  TRAVEL.XLS.

Город	Гостиница	Прокат автомобиля
Сиэтл	176	45
Сан-Франциско	178	42
Лос-Анджелес	223	36
Феникс	124	38
Денвер	139	38
Миннеаполис	167	53
Чикаго	257	51
Сент-Луис	159	53
Даллас	167	46
Хьюстон	180	48
Детройт	141	53
Кливленд	145	40
Новый Орлеан	142	49
Питсбург	148	49
Атланта	173	46
Бостон	243	46
Нью-Йорк	273	69
Вашингтон	262	47
Орlando	133	40
Майами	116	39

Источник: “Travel”, The Wall Street Journal, 7 июня, 2002.

1. Постройте интервал, содержащий среднюю стоимость номера в гостинице, доверительный уровень которого равен 95%.
2. Постройте интервал, содержащий среднюю стоимость проката автомобиля, доверительный уровень которого равен 95%.
3. Какие ограничения следует наложить на распределение интересующих нас генеральных совокупностей при решении задач 1 и 2?
4. Выполняются ли предположения, необходимые для решения задач 1 и 2? Аргументируйте свой ответ.

7.20. Пресс разрезает куски стали на части, которые в дальнейшем используются в качестве каркаса переднего сиденья автомобиля. Для разрезания стали используется алмазная пила. Автомобильная компания постановила, что отклонение объемов каркаса от эталона не должно превышать 0,005 дюйма. В файле  STEEL.XLS при-

ведены отклонения от эталона объемов 100 заготовок, измеренных с помощью лазерных приборов. Например, величина $-0,002$ означает, что заготовка короче эталона на $0,002$ дюйма.

1. Постройте 95%-ный доверительный интервал, содержащий среднюю разность между фактическим и эталонным объемом каркаса.
2. Можно ли утверждать, что предположения, необходимые для решения задачи 1, не выполняются? Аргументируйте свой ответ.
3. Сравните выводы, полученные при решении задачи 1, с результатами решения задачи 2.16.

7.21. В одной из статей, опубликованных в журнале *Quality Engineering*, исследуется вязкость (т.е. величина сопротивления потоку) химического вещества из разных партий. В файле  CHEMICAL.XLS приведены данные о 120 партиях.

Источник: D. S. Holmes, and Mergen A. E., "Parabolic Control Limits for the Exponentially Weighted Moving Average Control Charts", Quality Engineering, 4(1992): p. 487–495.

1. Постройте 99%-ный доверительный интервал, содержащий среднюю вязкость вещества.
2. Какие ограничения следует наложить на распределение интересующей нас генеральной совокупности при решении задачи 1?
3. Можно ли утверждать, что предположения, необходимые для решения задачи 1, не выполняются? Аргументируйте свой ответ.
4. Сравните выводы, полученные при решении задачи 1, с результатами решения задачи 2.64.

7.3. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ ДОЛИ ПРИЗНАКА В ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

В этом разделе понятие доверительного интервала распространяется на категориальные данные. Это позволяет оценить долю признака в генеральной совокупности p с помощью выборочной доли $p_s = X/n$. Как указывалось в разделе 6.7, если величины np и $n(1-p)$ превышают число 5, биномиальное распределение можно аппроксимировать нормальным. Следовательно, для оценки доли признака в генеральной совокупности p можно построить интервал, доверительный уровень которого равен $(1-\alpha) \times 100\%$.

ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ ДОЛИ ПРИЗНАКА В ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

$$p_s \pm Z \sqrt{\frac{p_s(1-p_s)}{n}},$$

или

$$p_s - Z \sqrt{\frac{p_s(1-p_s)}{n}} \leq p \leq p_s + Z \sqrt{\frac{p_s(1-p_s)}{n}}, \quad (7.3)$$

где p_s — выборочная доля признака, равная X/n , т.е. количеству успехов, деленному на объем выборки, p — доля признака в генеральной совокупности, Z — критическое значение стандартизованного нормального распределения, n — объем выборки.

7.3. Построение доверительного интервала для доли признака... 467

Чтобы продемонстрировать применение этой формулы, вернемся к задаче об аудиторской проверке в компании Saxon Home Improvement Company. Требуется определить частоту ошибок, нарушающих правила, установленные компанией. Для этого можно построить 95%-доверительный интервал, содержащий долю ошибочных накладных в генеральной совокупности всех накладных компании Saxon Home Improvement Company. Предположим, что из информационной системы извлечена выборка, состоящая из 100 накладных, заполненных в течение последнего месяца. Допустим, что 10 из этих накладных составлены с ошибками. Таким образом, $p_s = 10/100 = 0,10$. Доверительному уровню 95% соответствует критическое значение $Z = 1,96$.

$$p_s \pm Z \sqrt{\frac{p_s(1-p_s)}{n}} = 0,10 \pm 1,96 \sqrt{\frac{0,10 \times 0,90}{100}} = 0,10 \pm 1,96 \times 0,03 = 0,10 \pm 0,0588.$$

$$0,0412 \leq p \leq 0,1588.$$

Таким образом, вероятность того, что от 4,12 до 15,88% накладных содержат ошибки, равна 95%. На рис. 7.10 продемонстрировано вычисление доверительного интервала, содержащего долю ошибок в генеральной совокупности накладных, с помощью программы Microsoft Excel.

	A	B
1	Доля ошибочных накладных	
2		
3	Data	
4	Sample Size	100
5	Number of Successes	10
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Sample Proportion	0.1
10	Z Value	-1.95996279
11	Standard Error of the Proportion	0.03
12	Interval Half Width	0.058798884
13		
14	Confidence Interval	
15	Interval Lower Limit	0.041201116
16	Interval Upper Limit	0.158798884

Рис. 7.10. Вычисление доверительного интервала, содержащего долю ошибок в генеральной совокупности накладных, с помощью программы Microsoft Excel

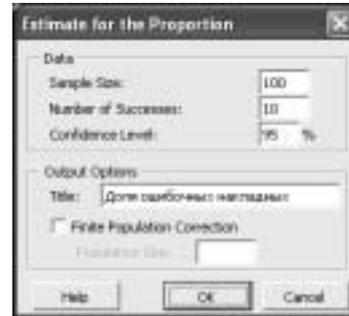
Процедуры Excel: построение доверительного интервала для доли признака в генеральной совокупности ▼

Чтобы построить доверительный интервал для доли признака в генеральной совокупности, можно воспользоваться функцией `НОРМСТОБР` или надстройкой `PHStat2`. Например, чтобы построить доверительный интервал, показанный на рис. 7.10, необходимо выполнить одну из следующих процедур.

Применение Excel в сочетании с надстройкой PHStat2

Чтобы построить доверительный интервал для доли признака в генеральной совокупности при неизвестном стандартном отклонении, следует применить процедуру `PHStat⇒Confidence Intervals⇒Estimate for the Proportion...` (`PHStat⇒Доверительные интервалы⇒Оценка доли...`). Для этого необходимо выполнить инструкции, приведенные ниже.

1. Выбрать команду PHStat⇒Confidence Intervals⇒Estimate for the Proportion...
2. В диалоговом окне Estimate for the Proportion (см. иллюстрацию) сделать следующее.
 - 2.1. Ввести в окне редактирования Sample Size (Объем выборки) число 100.
 - 2.2. Ввести в окне редактирования Number of Success (Количество успехов) число 10.
 - 2.3. Ввести в окне редактирования Confidence Level (Доверительный интервал) число 95.
 - 2.4. Ввести в окне редактирования Title соответствующий заголовок.
 - 2.5. Щелкнуть на кнопке ОК.



Применение Excel

Для того чтобы самостоятельно создать рабочий лист, вычисляющий доверительный интервал для доли признака в генеральной совокупности при неизвестном стандартном отклонении, следуйте инструкциям, приведенным в разделе ЕН.7.3.

Chapter 7.xls

Данные, на основе которых вычисляется доверительный интервал, показанный на рис. 7.10, содержатся в рабочей книге Chapter 7.xls на листе Рис7.10.

ПРИМЕР 7.4. ОЦЕНКА ДОЛИ БРАКОВАННЫХ ЭКЗЕМПЛЯРОВ ГАЗЕТЫ

Редактор крупной газеты желает оценить среднегодовую долю бракованных экземпляров в тираже. Бракованными считаются экземпляры, содержащие пятна, нарушение нумерации страниц, пропуски и дубликаты страниц. Для этого он создал выборку, состоящую из 200 экземпляров. В данной выборке 35 экземпляров оказались бракованными. Постройте интервал, содержащий долю брака с 90%-ным доверительным уровнем.

РЕШЕНИЕ. Доверительный интервал вычисляется следующим образом. Доверительному уровню 90% соответствует критическое значение $Z = 1,645$.

$$p_s = 35 / 200 = 0,175,$$

$$p_s \pm Z \sqrt{\frac{p_s(1-p_s)}{n}} = 0,175 \pm 1,645 \sqrt{\frac{0,175 \times 0,825}{200}} = 0,175 \pm 1,645 \times 0,0269 = 0,175 \pm 0,0442,$$

$$0,1308 \leq p \leq 0,2192.$$

Таким образом, вероятность того, что от 13,08 до 21,92% экземпляров окажутся бракованными, равна 90%. ■

ПРОВЕРКА ПРЕДПОЛОЖЕНИЙ

В примере 7.4 количество успехов и неудач относительно велико, поэтому биномиальное распределение можно аппроксимировать нормальным. Однако, если объем выборки невелик или количество успехов мало, следует использовать биномиальное, а не нормальное распределение [1, 6]. Таблицы, содержащие точные доверительные интервалы для разных объемов выборок и долей успеха, созданы Фишером и Йейтсом (Fisher and Yates) [2].

Для заданного объема выборки доверительный интервал, содержащий долю признака в генеральной совокупности, кажется более широким, чем для непрерывной случайной величины. Это объясняется тем, что измерения непрерывной случайной величины содержат больше информации, чем измерения категориальных данных. Иначе говоря, категориальные данные, принимающие лишь два значения, содержат недостаточно информации для оценки параметров их распределения.

УПРАЖНЕНИЯ К РАЗДЕЛУ 7.3

Изучение основ

- 7.22.** Вычислите 95%-ный доверительный интервал, содержащий долю признака в генеральной совокупности, если $n = 200$, а $X = 50$.
- 7.23.** Вычислите 99%-ный доверительный интервал, содержащий долю признака в генеральной совокупности, если $n = 400$, а $X = 25$.

Применение понятий

- 7.24.** Телефонной компании необходимо оценить количество владельцев телефонных номеров, желающих установить дополнительную телефонную линию по сниженной цене. Для этого была создана случайная выборка, состоящая из 500 клиентов. Оказалось, что 135 владельцев телефонных номеров согласны оплатить эту услугу.
1. Постройте интервал, содержащий долю владельцев телефонных номеров, желающих установить дополнительную телефонную линию, доверительный уровень которого равен 99%.
 2. Какие выводы может сделать менеджер телефонной компании, анализируя результаты решения задачи 1?
- 7.25.** В статье, недавно опубликованной в журнале *Wall Street Journal*, утверждается, что женщины начинают новое дело, стремясь к свободе и независимости, а не из-за денег (“Work Week”, *Wall Street Journal*, June 8, 1999, p. A1). В опросе участвовали 763 женщины, основавшие новые компании, 229 из них ответили, что начали новое дело, стремясь к большей свободе. Только 99 женщин заявили, что занимаются бизнесом, чтобы заработать больше денег.
1. Постройте интервал с 90%-ным доверительным уровнем, содержащий долю женщин, которые начали новое дело, стремясь к свободе и независимости.
 2. Постройте интервал с 90%-ным доверительным уровнем, содержащий долю женщин, которые начали новое дело, желая заработать больше денег.
- 7.26.** Исследование, проведенное организацией Randstad North America, показало, что только 41% из 1 110 американских рабочих считают, что их начальники относятся к ним лояльно (Carlos Tejada, “Work Week”, *Wall Street Journal*, May 29, 2002, B10).
1. Постройте 95%-ный доверительный интервал, содержащий долю рабочих, считающих, что их начальники относятся к ним лояльно.
 2. Постройте 90%-ный доверительный интервал, содержащий долю рабочих, считающих, что их начальники относятся к ним лояльно.
 3. Какой из этих интервалов шире? Аргументируйте свой ответ.
- 7.27.** В ходе исследования, проведенного журналом *Reader's Digest* (“Testing Honesty Around the World”, *USA Today*, May 1, 2001, 1A), в разных странах мира были “потеряны” 1 100 кошельков, в которых лежали 50 долл. и записка с фамилией и телефонным номером “хозяина”. В 484 случаях люди, нашедшие кошелек, не вернули его “хозяину”. Постройте 95%-ный доверительный интервал для доли “потерянных” кошельков, возвращенных “владельцам”.

- 7.28.** Опрос автомобилистов, использующих мобильные телефоны, показал, что 46% респондентов сбились с пути, а 10% заявили, что знают случаи, когда люди попадали в аварию, разговаривая по мобильному телефону (“Drivers Using Cell Phones Have Problems”, *USA Today*, May 16, 2001, 1A). Предположим, что в исследовании приняли участие 500 респондентов.
1. Постройте 95%-ный доверительный интервал, содержащий долю автомобилистов, которые сбились с пути, разговаривая по мобильному телефону.
 2. Постройте 95%-ный доверительный интервал, содержащий долю автомобилистов, знающих о случаях, когда люди попадали в аварию, разговаривая по мобильному телефону.
- 7.29.** Компания Walker Information, Inc. спросила у 2 800 американских рабочих, испытывают ли они чувство неловкости, сообщая руководству о плохой работе своих коллег. В 1 456 случаях респонденты ответили утвердительно, а 1 344 рабочих ответили отрицательно (“Work Week”, *Wall Street Journal*, September 4, 2001, A1).
1. Постройте 95%-ный доверительный интервал, содержащий долю американских рабочих, которые испытывают чувство неловкости, сообщая руководству о плохой работе своих коллег.
 2. Постройте 99%-ный доверительный интервал, содержащий долю американских рабочих, которые испытывают чувство неловкости, сообщая руководству о плохой работе своих коллег.
 3. Уверены ли вы, что более половины американских рабочих испытывают чувство неловкости, сообщая руководству о плохой работе своих коллег?
 4. В статье утверждается также, что в 1999 году 34% всех американских рабочих не испытывали чувства неловкости, сообщая руководству о плохой работе своих коллег. Уверены ли вы, что этот процент значительно увеличился с 1999 по 2001 гг.?
- 7.30.** Клиническое подразделение компании Estee Lauder Cosmetics провело в Северной Америке опрос работающих женщин. Из 1 000 опрошенных женщин 55% полагают, что компании должны резервировать место работы для женщин, находящихся в декретном отпуске, на срок до 6 месяцев, а остальные 45% считают, что этот срок должен превышать 6 месяцев (“Work Week”, *Wall Street Journal*, September 11, 2001, A1).
1. Постройте 95%-ный доверительный интервал, содержащий долю североамериканских женщин, полагающих, что компании должны резервировать место работы для женщин, находящихся в декретном отпуске, на срок более 6 месяцев.
 2. Дайте интерпретацию доверительного интервала, построенного при решении задачи 1.
- 7.31.** На протяжении июня и июля 2001 года комиссия Европейского Союза провела опрос 6 543 взрослых европейцев. Среди опрошенных 56% заявили, что введение единой европейской валюты будет способствовать экономическому росту, а 73% точно знали дату ее введения (1 января 2002 года) (“Retail Cheating Seen as Likely”, *Marketing News*, September 24, 2001, 42).
1. Постройте 95%-ный доверительный интервал, содержащий долю взрослых европейцев, уверенных, что введение единой европейской валюты будет способствовать экономическому росту.
 2. Постройте 95%-ный доверительный интервал, содержащий долю взрослых европейцев, знающих точную дату введения единой европейской валюты.
-

7.4. ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ

В каждом из рассмотренных примеров мы заранее фиксировали объем выборки, не учитывая ширину доверительного интервала. В реальных задачах определить объем выборки довольно сложно. Это зависит от наличия финансовых ресурсов, времени и легкости создания выборки. Если нам необходимо оценить среднюю сумму накладных или долю ошибочных накладных в информационной системе компании *Saxon Home Improvement Company*, сначала следует выяснить, насколько точной должна быть оценка. Иначе говоря, следует задать ошибку выборочного исследования, допускаемую при оценке каждого из параметров. Кроме того, необходимо заранее определить доверительный уровень оценки истинного параметра генеральной совокупности.

Определение объема выборки для оценки математического ожидания

Чтобы определить объем выборки, необходимый для оценки математического ожидания генеральной совокупности, следует учесть величину ошибки выборочного исследования и доверительный уровень. Кроме того, необходима дополнительная информация о величине стандартного отклонения.

Для того чтобы вывести формулу, позволяющую вычислить объем выборки, напомним формулу (7.1):

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}},$$

где Z — критическое значение случайной величины, имеющей стандартизированное нормальное распределение.

Величина, добавляемая и вычитаемая из \bar{X} , равна половине длины интервала. Она определяет меру неточности оценки, возникающей вследствие **ошибки выборочного исследования**, которая обозначается символом e и вычисляется по формуле

$$e = Z \frac{\sigma}{\sqrt{n}}.$$

Следовательно, объем выборки n определяется по такой формуле.

ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ ДЛЯ ОЦЕНКИ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ

Объем выборки равен произведению квадрата величины Z на дисперсию σ^2 , деленному на квадрат ошибки выборочного исследования e :

$$n = \frac{Z^2 \sigma^2}{e^2}. \quad (7.4)$$

Таким образом, для определения объема выборки необходимо знать три параметра.

1. Требуемый доверительный уровень, который влияет на величину Z , являющуюся критическим значением стандартизированного нормального распределения.³
2. Приемлемую ошибку выборочного исследования e ;
3. Стандартное отклонение σ .

³ Для определения размера выборки используется величина Z , а не t , поскольку для вычисления критического значения t размер выборки необходимо знать заранее. В большинстве случаев размеры выборки позволяют хорошо аппроксимировать t -распределение стандартизированным нормальным распределением.

На практике вычислить эти величины непросто. Как определить доверительный уровень и ошибку выборочного исследования? Обычно ответить на этот вопрос могут лишь эксперты в предметной области (т.е. люди, понимающие смысл оцениваемых величин). Как правило, доверительный уровень равен 95% (в этом случае $Z = 1,96$). Если требуется поднять доверительный уровень, обычно выбирают величину, равную 99%. Если можно ограничиться более низким доверительным уровнем, выбирают 90%. Определяя ошибку выборочного исследования, не стоит думать о ее величине (в принципе, любая ошибка нежелательна). Следует задать такую ошибку, чтобы полученные результаты допускали разумную интерпретацию.

Кроме доверительного уровня и ошибки выборочного исследования, необходимо знать стандартное отклонение генеральной совокупности. К сожалению, этот параметр почти никогда не известен. В некоторых случаях стандартное отклонение генеральной совокупности можно оценить на основе предшествующих исследований. В других ситуациях опытный эксперт может учесть размах выборки и распределение случайной переменной. Например, если генеральная совокупность имеет нормальное распределение, ее размах приближенно равен 6σ (т.е. $\pm 3\sigma$ в окрестности математического ожидания). Следовательно, стандартное отклонение приближенно равно одной шестой части диапазона. Если величину σ невозможно оценить таким способом, необходимо выполнить пилотный проект и вычислить стандартное отклонение по результатам.

Чтобы продемонстрировать применение этого подхода, вернемся к задаче об аудиторской проверке в компании *Saxon Home Improvement Company*. Предположим, что из информационной системы извлечена выборка, состоящая из 100 накладных, заполненных в течение последнего месяца. Компания желает построить интервал, содержащий математическое ожидание генеральной совокупности, доверительный уровень которого равен 95%. Как был определен объем выборки? Следует ли его уточнить?

Допустим, что после консультаций с экспертами, работающими в компании, статистики установили допустимую ошибку выборочного исследования равной ± 5 долл., а доверительный уровень — 95%. Результаты предшествующих исследований свидетельствуют, что стандартное отклонение генеральной совокупности приближенно равно 25 долл. Таким образом, $e = 5$, $\sigma = 25$ и $Z = 1,96$ (что соответствует 95%-ному доверительному уровню). По формуле (7.4) получаем:

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1,96)^2 25^2}{5^2} = 96,04.$$

Следовательно, $n = 97$, поскольку дробные результаты, как правило, округляют с избытком до ближайшего целого. Таким образом, объем выборки, равный 100, был выбран удачно и вполне соответствует требованиям, выдвинутым компанией. Однако, если стандартное отклонение равно 28,95, доверительный интервал оказывается немного шире, чем требуется. Пример вычисления объема выборки с помощью программы *Microsoft Excel* приведен на рис. 7.11.

	A	B
1	Для средней суммы накладных	
2		
3	Data	
4	Population Standard Deviation	25
5	Sampling Error	5
6	Confidence Level	95%
7	Intermediate Calculations	
8	Z Value	-1.95996279
9	Calculated Sample Size	96.036302
10		
11	Result	
12	Sample Size Needed	97

Рис. 7.11. Вычисление объема выборки при построении доверительного интервала для математического ожидания генеральной совокупности накладных с помощью программы *Microsoft Excel*

Процедуры Excel: определение объема выборки при оценке математического ожидания генеральной совокупности ▼

Чтобы определить объем выборки, необходимой для построения доверительного интервала, содержащего математическое ожидание генеральной совокупности, можно воспользоваться функцией НОРМСТОВБР или надстройкой PHStat2.

Например, чтобы определить объем выборки для оценки средней суммы накладных компании Saxon Home Improvement Company, необходимо выполнить одну из следующих процедур.

Применение Excel в сочетании с надстройкой PHStat2

Чтобы определить объем выборки, необходимой для построения доверительного интервала, содержащего математическое ожидание генеральной совокупности, следует применить процедуру PHStat⇒Sample Size⇒Determination for the Mean... (PHStat⇒Объем выборки⇒Определение математического ожидания...).

1. Выбрать PHStat⇒Sample Size⇒Determination for the Mean...
2. В диалоговом окне Sample Size Determination for the Mean Estimate (Определение объема выборки для оценки математического ожидания) (см. иллюстрацию) сделать следующее.
 - 2.1. Ввести в окне редактирования Population Standard Deviation (Стандартное отклонение генеральной совокупности) число 25.
 - 2.2. Ввести в окне редактирования Sampling Error (Ошибка выборочного исследования) число 5.
 - 2.3. Ввести в окне редактирования Confidence Level (Доверительный уровень) число 95.
 - 2.4. Ввести в окне редактирования Title соответствующий заголовок.
 - 2.5. Щелкнуть на кнопке ОК.



Применение Excel

Для того чтобы самостоятельно создать рабочий лист, определяющий объем выборки, необходимый для вычисления доверительного интервала, содержащего математическое ожидание генеральной совокупности, следуйте инструкциям, приведенным в разделе ЕН.7.4.



Chapter 7.xls

Данные, на основе которых определяется объем выборки, необходимый для вычисления доверительного интервала, содержащего математическое ожидание генеральной совокупности накладных, содержатся в рабочей книге Chapter 7.xls на листе Рис7.11.

ПРИМЕР 7.5. ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ ДЛЯ ОЦЕНКИ СРЕДНЕГО

Вернемся к примеру 7.3. Предположим, что нам необходимо оценить среднюю силу разрушения изолятора с точностью ± 25 фунтов и построить 95%-ный доверительный интервал для этой величины. Данные, полученные в предыдущем исследовании, свидетельствуют, что стандартное отклонение равно 100 фунтов. Определите требуемый объем выборки.

РЕШЕНИЕ. Итак, $e = 25$, $\sigma = 100$, а доверительный уровень равен 95% (т.е. $Z = 1,96$).

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1,96)^2 100^2}{25^2} = 61,47.$$

Таким образом, $n = 62$. Напомним, что дробные результаты, как правило, округляют с избытком до ближайшего целого.

Обратите внимание на то, что при указанных данных объем выборки равен 62, а не 30, как было установлено в примере 7.3. Кроме того, величина стандартного отклонения определена на основе прошлогодних наблюдений. Если реальное стандартное отклонение значительно отличается от этой величины, ошибка выборочного исследования окажется совершенно иной. ■

Определение объема выборки для оценки доли признака в генеральной совокупности

Выше мы рассмотрели способ определения объема выборки для оценки математического ожидания генеральной совокупности. Предположим теперь, что мы проводим аудиторскую проверку в компании Saxon Home Improvement Company. Нам необходимо определить долю накладных, не соответствующих правилам, принятым компанией. Сколько накладных следует извлечь из информационной системы, чтобы построенный интервал имел заданный доверительный уровень? Для ответа на этот вопрос применим тот же подход, что и при определении объема выборки для оценки математического ожидания.

Напомним, что ошибка выборочного исследования определяется по формуле:

$$e = Z \frac{\sigma}{\sqrt{n}}.$$

При оценке доли признака величину σ следует заменить на величину $\sqrt{p(1-p)}$. Таким образом, формула для ошибки выборочного исследования принимает следующий вид:

$$e = Z \sqrt{\frac{p(1-p)}{n}}.$$

Выражая n через остальные величины, получаем следующую формулу.

ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ ДЛЯ ОЦЕНКИ ДОЛИ ПРИЗНАКА

Объем выборки равен произведению квадрата величины Z на $p(1-p)$, деленному на квадрат ошибки выборочного исследования e :

$$n = \frac{Z^2 p(1-p)}{e^2}. \quad (7.5)$$

Таким образом, для определения объема выборки необходимо знать три параметра.

1. Требуемый доверительный уровень, по которому определяется величина Z .
2. Допустимую ошибку выборочного исследования e .
3. Истинную долю успехов p .

На практике вычислить эти величины нелегко. Если известен доверительный уровень, можно вычислить критическое значение стандартизованного нормального распределения Z . Ошибка выборочного исследования e определяет точность, с которой оценивается доля успехов в генеральной совокупности. Третий параметр — доля успехов в генеральной совокупности p — это именно тот параметр, который нам необходимо оценить. Итак, как оценить диапазон изменения величины p по его выборочным значениям?

Существуют два способа. Во-первых, во многих ситуациях для оценки величины p можно использовать результаты предыдущих исследований. Во-вторых, если данные

о предыдущих исследованиях недоступны, можно попытаться оценить параметр p так, чтобы исключить недооценку объема выборки. Обратите внимание на то, что в формуле (7.5) величина $p(1-p)$ стоит в числителе. Следовательно, необходимо найти максимальное значение этой величины. Очевидно, что оно достигается при $p = 0,5$. Перечислим некоторые значения произведения $p(1-p)$.

Если $p = 0,9$, то $p(1-p) = 0,9 \times 0,1 = 0,09$.

Если $p = 0,7$, то $p(1-p) = 0,7 \times 0,3 = 0,21$.

Если $p = 0,5$, то $p(1-p) = 0,5 \times 0,5 = 0,25$.

Если $p = 0,3$, то $p(1-p) = 0,3 \times 0,7 = 0,21$.

Если $p = 0,1$, то $p(1-p) = 0,1 \times 0,9 = 0,09$.

Таким образом, если доля признака в генеральной совокупности p заранее неизвестна, для определения объема выборки следует задать $p = 0,5$. В этом случае объем выборки будет переоценен, что приведет к дополнительным затратам на ее создание. Если истинная доля успехов в генеральной совокупности сильно отличается от 0,5, доверительный интервал окажется значительно уже, чем требовалось. Оценка параметра p в этом случае будет весьма точной, однако за это придется заплатить дополнительными временными и финансовыми ресурсами.

Вернемся к задаче об аудиторской проверке в компании Saxon Home Improvement Company. Предположим, аудитор желает построить интервал, содержащий долю ошибочных накладных, доверительный уровень которого равен 95%. Допустимая точность равна $\pm 0,07$. Результаты предыдущих проверок свидетельствуют, что доля ошибочных накладных не превышает 0,15. Таким образом, $e = 0,07$, $p = 0,15$ и $Z = 1,96$ (что соответствует 95%-ному доверительному уровню). По формуле (7.5) получаем:

$$n = \frac{Z^2 p(1-p)}{e^2} = \frac{(1,96)^2 \times 0,15 \times 0,85}{(0,07)^2} = 99,96.$$

Следовательно, $n = 100$, поскольку дробные результаты, как правило, округляют с избытком до ближайшего целого. Таким образом, объем выборки, равный 100, был выбран совершенно правильно и вполне соответствует требованиям, выдвинутым компанией. Однако, если стандартное отклонение равно 0,10, доверительный интервал оказывается немного уже, чем требуется. Пример вычисления объема выборки с помощью программ PHStat и Microsoft Excel приведен на рис. 7.12.

	A	B
1	Для доли ошибочных накладных	
2		
3	Data	
4	Estimate of True Proportion	0.15
5	Sampling Error	0.07
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Z Value	-1.95996279
10	Calculated Sample Size	99.99620435
11		
12	Result	
13	Sample Size Needed	100

Рис. 7.12. Вычисление объема выборки при построении доверительного интервала, содержащего долю признака в генеральной совокупности накладных, с помощью программы Microsoft Excel

Процедуры Excel: определение объема выборки при оценке доли признака в генеральной совокупности

Чтобы определить объем выборки, необходимой для построения доверительного интервала, содержащего долю признака в генеральной совокупности, можно воспользоваться функцией `НОРМСТОБР` или надстройкой PHStat2.

Например, чтобы определить объем выборки для оценки доли ошибочных накладных, как показано на рис. 7.12, необходимо выполнить одну из следующих процедур.

Применение Excel в сочетании с надстройкой PHStat2

Чтобы определить объем выборки, необходимой для построения доверительного интервала, содержащего долю признака в генеральной совокупности, следует применить процедуру PHStat2→Sample Size→Determination for the Proportion... (PHStat2→Объем выборки→Определение для доли признака...).

1. Выбрать PHStat2→Sample Size→Determination for the Proportion...
2. В диалоговом окне Sample Size Determination for the Proportion (Определение объема выборки для оценки доли признака) (см. иллюстрацию) сделать следующее.
 - 2.1. Ввести в окне редактирования Estimate of True Proportion (Оценка истинной доли) число 0.15.
 - 2.2. Ввести в окне редактирования Sampling Error (Ошибка выборочного исследования) число 0.07.
 - 2.3. Ввести в окне редактирования Confidence Level (Доверительный уровень) число 95.
 - 2.4. Ввести в окне редактирования Title соответствующий заголовок.
 - 2.5. Щелкнуть на кнопке ОК.



Применение Excel

Для того чтобы самостоятельно создать рабочий лист, определяющий объем выборки, необходимый для вычисления доверительного интервала, содержащего долю признака в генеральной совокупности, следуйте инструкциям, приведенным в разделе ЕН.7.5.

 Chapter 7.xls

Данные, на основе которых определяется объем выборки для вычислений, показанных на рис. 7.12, содержатся в рабочей книге Chapter 7.xls на листе Рис7.12.

ПРИМЕР 7.6. ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ ДЛЯ ОЦЕНКИ ДОЛИ ПРИЗНАКА В ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Редактор крупной газеты, упомянутой в примере 7.4, желает построить интервал, содержащий долю брака с 90%-ным доверительным уровнем. Допустимая ошибка выборочной проверки равна $\pm 0,05$. Кроме того, редактору неизвестны результаты предшествующих проверок. Определите необходимый объем выборки.

РЕШЕНИЕ. Поскольку никаких данных о предыдущих проверках нет, примем $p = 0,50$. Кроме того, $e = 0,05$, а доверительный уровень равен 90% (т.е. $Z = 1,645$).

$$n = \frac{(1,645)^2 \times 0,50 \times 0,50}{(0,05)^2} = 270,6.$$

Таким образом, $n = 271$. Итак, чтобы построить интервал, содержащий долю брака с 90%-ным доверительным уровнем при допустимой ошибке выборочного исследования, равной $\pm 0,05$, выборка должна содержать 271 экземпляр газеты. ■

УПРАЖНЕНИЯ К РАЗДЕЛУ 7.4

Изучение основ

- 7.32.** Вычислите объем выборки, необходимой для построения интервала, содержащего математическое ожидание генеральной совокупности с 95%-ным доверительным уровнем, если допустимая ошибка выборочного исследования равна ± 5 , а стандартное отклонение — 15.
- 7.33.** Вычислите объем выборки, необходимой для построения интервала, содержащего математическое ожидание генеральной совокупности с 99%-ным доверительным уровнем, если допустимая ошибка выборочного исследования равна ± 20 , а стандартное отклонение — 100.
- 7.34.** Вычислите объем выборки, необходимой для построения интервала, содержащего долю признака в генеральной совокупности с 99%-ным доверительным уровнем, если допустимая ошибка выборочного исследования равна $\pm 0,04$.
- 7.35.** Вычислите объем выборки, необходимой для построения интервала, содержащего долю признака в генеральной совокупности с 99%-ным доверительным уровнем, если допустимая ошибка выборочного исследования равна $\pm 0,04$, а предыдущие исследования показали, что приблизительное значение параметра p равно 0,40.

Применение понятий

- 7.36.** В крупной компании планируется опрос сотрудников. Цель опроса — выяснить среднегодовую величину расходов на медицинские услуги. Администрация компании желает построить 95%-ный доверительный интервал, содержащий математическое ожидание генеральной совокупности с точностью ± 50 долл. Предварительный опрос показал, что стандартное отклонение приблизительно равно 400 долл.
1. Определите необходимый объем выборки.
 2. Предположим, что администрация компании решила повысить точность прогноза до ± 25 долл. Определите необходимый объем выборки.
- 7.37.** Директор магазина стройматериалов желает оценить средний объем краски, содержащейся в банке, емкостью один галлон с точностью $\pm 0,004$. Для этого он собирается построить 95%-ный доверительный интервал, предполагая, что стандартное отклонение равно 0,02 галлона. Определите необходимый объем выборки.
- 7.38.** Инспектор по контролю за качеством продукции желает оценить среднюю продолжительность работы электрических лампочек с точностью ± 20 . Для этого он собирается построить 95%-ный доверительный интервал, предполагая, что стандартное отклонение равно 100 ч. Определите необходимый объем выборки.
- 7.39.** Метролог компании, производящей безалкогольные напитки, желает оценить средний объем жидкости в двухлитровых бутылках с точностью $\pm 0,01$ л. Для этого он собирается построить 95%-ный доверительный интервал, предполагая, что стандартное отклонение равно 0,05 л. Определите объем выборки.
- 7.40.** Общество защиты прав потребителей желает оценить среднюю сумму, затраченную обычной семьей на оплату электричества за июль. Предварительные исследования, проведенные в других городах, показали, что стандартное отклонение равно 25 долл. Допустимая точность оценки равна ± 5 долл. Необходимо построить 99%-ный доверительный интервал, содержащий среднюю сумму затрат на электричество.

1. Определите необходимый объем выборки.
 2. Предположим, что доверительный уровень равен 95%. Определите необходимый объем выборки.
- 7.41.** Рекламное агентство, заключившее договор с местной радиостудией, желает оценить среднее время, которое слушатели ежедневно проводят перед радиоприемником. Предварительные исследования показали, что стандартное отклонение равно 45 мин.
1. Определите объем выборки, если допустимая точность оценки равна ± 5 мин., а доверительный уровень — 90%.
 2. Предположим, что доверительный уровень равен 99%. Определите необходимый объем выборки.
- 7.42.** Компания, поставляющая газ, желает оценить среднее время ожидания своих потребителей. Определите необходимый объем выборки, если допустимая точность оценки равна ± 5 дней, доверительный уровень — 95%, а стандартное отклонение равно 20 дней.
- 7.43.** Файл  NYSE.XLS содержит случайную выборку, состоящую из стоимости 18 акций, котируемых на Нью-Йоркской фондовой бирже по состоянию на 4 марта 2003 года.
1. Постройте 95%-ный доверительный интервал, содержащий математическое ожидание этой генеральной совокупности.
 2. Какой объем должна иметь выборка, если доверительный уровень равен 95%, а допустимая точность — $\pm 20\,000$ акций?
 3. Сравните ваши оценки с реальными данными. Обоснуйте свои рассуждения.
- 7.44.** Социологическая служба желает оценить долю избирателей, собирающихся проголосовать на предстоящих президентских выборах за представителя демократической партии. Необходимо построить 90%-ный доверительный интервал с точностью $\pm 0,04$.
1. Определите необходимый объем выборки.
 2. Предположим, что доверительный уровень равен 95%. Определите необходимый объем выборки.
 3. Предположим, что доверительный уровень равен 95%, а точность — $\pm 0,03$. Определите необходимый объем выборки.
 4. Объясните, как объем выборки зависит от доверительного уровня и ошибки выборочного исследования. Аргументируйте свой ответ.
- 7.45.** В 2001 г. 45% семей в США приобретали бакалейные товары в ближайших лавках, а 29% семей — в оптовых магазинах (“68th Annual Report of the Grocery Industry”, Progressive Grocer, April 2002, 29).
1. Определите объем выборки, необходимый для оценки доли американских семей, приобретающих бакалейные товары в ближайших лавках, если точность выборочного исследования равна $\pm 0,02$, а доверительный уровень — 95%.
 2. Определите объем выборки, необходимый для оценки доли американских семей, приобретающих бакалейные товары в оптовых магазинах, если точность выборочного исследования равна $\pm 0,02$, а доверительный уровень — 95%.
 3. Сравните результаты решения задач 1 и 2. Объясните разницу между ними.

4. Предположим, вы планируете дальнейшие исследования. Можно ли использовать одну и ту же выборку, задавая респондентам оба вопроса одновременно, или следует сформировать две отдельные выборки? Аргументируйте свой ответ.
 5. Как сформировать выборки для исследования в задаче 4?
- 7.46.** Сколько людей, живущих в США, планируют свой отпуск с помощью Интернет? Согласно опросу, проведенному компанией American Express, доля таких людей составляет 35% (A. R. Carey and K. Carter, "Snapshots", *USA Today*, January 14, 2003, 1A). Предположим, что вы планируете свое исследование.
1. Определите объем выборки, необходимый для оценки доли американцев, планирующих свой отпуск с помощью Интернет, если точность выборочного исследования равна $\pm 0,04$, а доверительный уровень — 95%.
 2. Определите объем выборки, необходимый для оценки доли американцев, планирующих свой отпуск с помощью Интернет, если точность выборочного исследования равна $\pm 0,04$, а доверительный уровень — 99%.
 3. Определите объем выборки, необходимый для оценки доли американцев, планирующих свой отпуск с помощью Интернет, если точность выборочного исследования равна $\pm 0,02$, а доверительный уровень — 95%.
 4. Определите объем выборки, необходимый для оценки доли американцев, планирующих свой отпуск с помощью Интернет, если точность выборочного исследования равна $\pm 0,02$, а доверительный уровень — 99%.
 5. Объясните, как изменение ошибки выборочного исследования и доверительного уровня влияют на требуемый объем выборки.
- 7.47.** Мешают ли звонки мобильных телефонов во время презентации деловых проектов? В ходе опроса 326 бизнесменов ответили "Да" и только 23 ответили "Нет". ("You say", *Presentations: Technology and Techniques for Effective Communication*, January 2003, 18.)
1. Постройте 95%-ный доверительный интервал для доли бизнесменов, считающих, что звонки мобильных телефонов мешают презентации деловых проектов.
 2. Дайте интерпретацию доверительного интервала, построенного при решении задачи 1.
 3. Предположим, что вы планируете собственное исследование. Определите объем выборки, необходимый для оценки доли бизнесменов, считающих, что звонки мобильных телефонов мешают презентации деловых проектов, если точность выборочного исследования равна $\pm 0,04$, а доверительный уровень — 95%.
 4. Предположим, что вы планируете собственное исследование. Определите объем выборки, необходимый для оценки доли бизнесменов, считающих, что звонки мобильных телефонов мешают презентации деловых проектов, если точность выборочного исследования равна $\pm 0,04$, а доверительный уровень — 99%.
 5. Предположим, что вы планируете собственное исследование. Определите объем выборки, необходимый для оценки доли бизнесменов, считающих, что звонки мобильных телефонов мешают презентации деловых проектов, если точность выборочного исследования равна $\pm 0,02$, а доверительный уровень — 95%.
 6. Предположим, что вы планируете собственное исследование. Определите объем выборки, необходимый для оценки доли бизнесменов, считающих, что звонки мобильных телефонов мешают презентации деловых проектов, если точность выборочного исследования равна $\pm 0,02$, а доверительный уровень — 99%.

7. Объясните, как изменение ошибки выборочного исследования и доверительного уровня влияют на требуемый объем выборки.
 8. Сравните свои выводы о влиянии ошибки выборочного исследования и доверительного уровня на требуемый объем выборки с решением задачи 7.46.
- 7.48.** Исследование, проведенное Федеральной резервной системой США, показало, что 52% из 4 449 семей в 2001 году владели акциями либо непосредственно, либо через взаимные фонды. (Barbara Hagenbaugh, “Nation’s Wealth Disparity Widens”, *USA Today*, January 22, 2003, 1A.)
1. Постройте 95%-ный доверительный интервал для доли семей, владевших акциями в 2001 году.
 2. Дайте интерпретацию доверительного интервала, построенного при решении задачи 1.
 3. Предположим, что вы планируете собственное исследование. Определите объем выборки, необходимый для оценки доли семей, владеющих акциями, если точность выборочного исследования равна $\pm 0,01$, а доверительный уровень — 95%.
- 7.49.** Исследование, проведенное фондом At-A-Glance Communications, показало, что 434 из 611 опрошенных клерков отвечают на полученное электронное письмо в течение одного или двух часов (D. Haralson and S. Ward, “You Have Mail”, *USA Today*, May 7, 2002, 1A.)
1. Постройте 95%-ный доверительный интервал для доли клерков, отвечающих на полученное электронное письмо в течение одного или двух часов
 2. Дайте интерпретацию доверительного интервала, построенного при решении задачи 1.
 3. Предположим, что вы планируете собственное исследование. Определите объем выборки, необходимый для оценки доли клерков, отвечающих на полученное электронное письмо в течение одного или двух часов, если точность выборочного исследования равна $\pm 0,01$, а доверительный уровень — 95%.

7.5. ПРИМЕНЕНИЕ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ В АУДИТОРСКОМ ДЕЛЕ

Описывая доверительные интервалы, мы сосредоточили внимание на математическом ожидании и доле признака в генеральной совокупности. Эти средства статистического анализа нашли весьма широкое применение в аудиторском деле.

Аудит — это сбор и оценка информации, позволяющей оценить состояние экономического объекта, например, компании, акционерного общества, корпорации или правительственного агентства. Цель аудита — оценить, насколько деятельность проверяемого объекта соответствует установленным критериям.

Во врезке 7.1 перечислены шесть основных преимуществ выборочного исследования, применяемого при аудите.

ВРЕЗКА 7.1. ПРЕИМУЩЕСТВА АУДИТОРСКОГО ОБСЛЕДОВАНИЯ

- Результаты выборочного исследования объективны и обоснованны. Поскольку определение объема выборки основано на точно сформулированных статистических принципах, результаты аудиторской проверки можно защищать в суде.
- Метод выборочного исследования позволяет заранее определить объем выборки.
- Метод позволяет оценить ошибку выборочного исследования.
- Этот подход можно применять для более точной оценки параметров, поскольку исследование большой генеральной совокупности может занять много времени и даже сопровождаться значительными ошибками нестатистического характера.
- Метод выборочного исследования могут применять сразу несколько аудиторов. Поскольку этот метод является научно обоснованным, можно считать, что в параллельной проверке принимает участие один аудитор.
- Метод выборочного исследования позволяет объективно оценить результаты проверки, поскольку его точность известна заранее.

Оценка суммы элементов генеральной совокупности

В аудиторском деле часто приходится оценивать сумму элементов генеральной совокупности. Она определяется по следующей формуле.

ОЦЕНКА СУММЫ ЭЛЕМЕНТОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Точечная оценка суммы элементов генеральной совокупности равна объему генеральной совокупности N , умноженному на выборочное среднее \bar{X} :

$$\text{Сумма} = N\bar{X}. \quad (7.6)$$

Границы доверительного интервала, содержащего сумму элементов генеральной совокупности, определяются по формуле (7.7).

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ, СОДЕРЖАЩИЙ СУММУ ЭЛЕМЕНТОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

$$N\bar{X} \pm N(t_{n-1}) \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}. \quad (7.7)$$

Чтобы продемонстрировать применение этой формулы, вернемся к задаче об аудиторской проверке в компании Saxon Home Improvement Company. Предположим, что для вычисления общей суммы накладных из информационной системы извлечены 100 записей, сделанных за последний месяц. Вся генеральная совокупность состоит из 5 000 накладных. Итак, $N = 5\,000$, $n = 100$, $\bar{X} = 110,27$, $S = 28,95$, доверительный уровень равен 95%, а $t_{99} = 1,9842$.

$$\text{Сумма} = 5\,000 \times 110,27 = 551\,350.$$

Таким образом, границы доверительного интервала равны:

$$\begin{aligned} N\bar{X} \pm N(t_{n-1}) \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} &= 551\,350 \pm 1,9842 \times 5\,000 \frac{28,95}{\sqrt{100}} \sqrt{\frac{5\,000-100}{5\,000-1}} = \\ &= 551\,350 \pm 28\,721,295 \times 0,99005. \\ 522\,914 &\leq \text{Сумма элементов генеральной совокупности} \leq 579\,786. \end{aligned}$$

Следовательно, вероятность того, что сумма элементов генеральной совокупности не меньше 522 914 и не больше 579 786, равна 95%. На рис. 7.13 продемонстрировано вычисление доверительного интервала, содержащего общую сумму генеральной совокупности накладных, с помощью программы Microsoft Excel.

	A	B
1	Общая сумма накладных	
2		
3	Data	
4	Population Size	5000
5	Sample Mean	110.27
6	Sample Size	100
7	Sample Standard Deviation	28.95
8	Confidence Level	95%
9		
10	Intermediate Calculations	
11	Population Total	551350.00
12	FPC Factor	0.9900489503
13	Standard Error of the Total	14330.95309
14	Degrees of Freedom	99
15	t Value	1.984217306
16	Interval Half Width	28435.72
17		
18	Confidence Interval	
19	Interval Lower Limit	522914.28
20	Interval Upper Limit	579785.72

Рис. 7.13. Вычисление доверительного интервала, содержащего общую сумму генеральной совокупности накладных, с помощью программы Microsoft Excel

Процедуры Excel: построение доверительного интервала для общей суммы генеральной совокупности

Чтобы построить доверительный интервал для общей суммы генеральной совокупности при неизвестном стандартном отклонении, можно воспользоваться функцией СТЬЮДРАСПОВР или надстройкой PHStat2.

Например, чтобы построить доверительный интервал для общей суммы накладных компании Saxon Home Improvement Company, показанный на рис. 7.13, необходимо открыть новый рабочий лист и выполнить одну из следующих процедур.

Применение Excel в сочетании с надстройкой Excel

Чтобы построить доверительный интервал для общей суммы генеральной совокупности, следует применить процедуру PHStat⇒Confidence Intervals⇒Estimate for the Population Total... (PHStat⇒Доверительные интервалы⇒Оценка общей суммы генеральной совокупности...).

1. Выбрать PHStat⇒Confidence Intervals⇒Estimate for the Population Total...
2. В диалоговом окне Estimate for the Population Total (см. иллюстрацию) сделать следующее.
 - 2.1. Ввести в окне редактирования Population Size (Объем выборки) число 5000.
 - 2.2. Ввести в окне редактирования Confidence Level (Доверительный интервал) число 95.
 - 2.3. Установить переключатель Input Options (Параметры выборки) в положение Sample Statistics Known (Выборочные статистики известны) и ввести в окне редактирования Sample Size (Объем выборки) число 100, в окне редактирования Sample Mean (Выборочное среднее) — число 110.27, а в окне редактирования Sample Std. Deviation (Выборочное стандартное отклонение) — число 28.95.



- 2.4. Ввести в окне редактирования Title соответствующий заголовок.
- 2.5. Щелкнуть на кнопке ОК.

Если выборочное среднее и выборочное стандартное отклонение не известны, в п. 2.2 следует установить переключатель Input Options (Параметры ввода) в положение Sample Statistics Unknown (Выборочные статистики известны), а в окне редактирования Sample Cell Range (Диапазон ячеек, содержащий выборку) ввести диапазон ячеек, в которых записаны элементы выборки.

Применение Excel

Для того чтобы самостоятельно создать рабочий лист, вычисляющий доверительный интервал, содержащий общую сумму элементов генеральной совокупности, следуйте инструкциям, приведенным в разделе ЕН.7.6.



Chapter 7.xls

Данные, на основе которых вычисляется доверительный интервал, содержащий общую сумму накладных, как показано на 7.13, хранятся в рабочей книге Chapter 7 на листе Рис7.13.

ПРИМЕР 7.7. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА, СОДЕРЖАЩЕГО ОБЩУЮ СУММУ ЭЛЕМЕНТОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Предположим, что в распоряжении аудитора оказалась генеральная совокупность чеков. Он желает оценить их общую сумму. Для этого аудитор извлекает из генеральной совокупности 50 чеков и вычисляет выборочное среднее и стандартное отклонение.

Выборочное среднее $\bar{X} = 1\,076,39$.

Стандартное отклонение $S = 273,62$ долл.

Необходимо построить 95%-ный доверительный интервал, содержащий общую сумму чеков.

РЕШЕНИЕ. Используя формулу (7.6), получим точечную оценку общей суммы чеков.

$$N\bar{X} = 1\,000 \times 1\,076,39 = 1\,076\,390.$$

Далее, по формуле (7.7) вычисляем границы 95%-ного доверительного интервала, содержащего общую сумму чеков.

$$\begin{aligned} N\bar{X} \pm N(t_{n-1}) \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} &= 1\,000 \times 1\,076,39 \pm 1\,000 \times 2,0096 \frac{273,62}{\sqrt{100}} \sqrt{\frac{1\,000-50}{1\,000-1}} = \\ &= 1\,076\,390 \pm 77\,762,902 \times 0,97517 = 1\,076\,390 \pm 75\,832. \\ 1\,000\,558 &\leq \text{Общая сумма чеков} \leq 1\,152\,222. \end{aligned}$$

Следовательно, вероятность того, что сумма элементов генеральной совокупности не меньше 1 000 558 и не больше 1 152 222, равна 95%. ■

Оценка разности

Оценка разности (difference estimation) применяется тогда, когда аудитор считает, что в анализируемой генеральной совокупности содержатся ошибки, которые необходимо оценить на основе выборочных данных. Для этого выполняются следующие процедуры.

1. Определяется необходимый объем выборки.
2. Вычисляются разности между величинами, обнаруженными в ходе аудита, и номинальными величинами. Обратите внимание на то, что разность D_i равна нулю, если обнаруженная величина правильна; является положительной величиной, если обнаруженная величина больше номинальной; и отрицательной — если обнаруженная величина меньше номинальной.
3. Вычисляется средняя выборочная разность \bar{D} , которая является результатом деления суммы разностей на объем выборки.

СРЕДНЯЯ РАЗНОСТЬ

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n}, \quad (7.8)$$

где D_i = обнаруженная величина – номинальная величина.

4. Вычисляется стандартное отклонение разностей S_D . Элементом выборки, не являющимся ошибкой, соответствуют нулевые разности.

СТАНДАРТНОЕ ОТКЛОНЕНИЕ РАЗНОСТЕЙ

$$S_D = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}}. \quad (7.9)$$

5. Вычисляются границы доверительного интервала, содержащего сумму разностей элементов генеральной совокупности.

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ, СОДЕРЖАЩИЙ ПОЛНУЮ РАЗНОСТЬ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

$$N\bar{D} \pm N(t_{n-1}) \frac{S_D}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}. \quad (7.10)$$

Продемонстрируем применение этой формулы на примере аудиторской проверки в компании Saxon Home Improvement Company. Предположим, что компания желает построить 95%-ный доверительный интервал, содержащий полную разность между фактическими суммами, указанными в накладных, и суммами, занесенными в информационную систему. В выборку, состоящую из 100 записей, входят 12 накладных, не соответствующих действительности. Перечислим эти разности, выраженные в долларах.



PLUMBINV.XLS

9,3 7,47 17,32 8,30 5,21 10,80 6,22 5,63 4,97 7,43 2,99 4,63

7.5. Применение доверительных интервалов в аудиторском деле 485

В остальных 88 накладных ошибок нет, следовательно, их *разности* равны нулю. Итак,

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n} = \frac{90}{100} = 0,90,$$

$$S_D = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}} = \sqrt{\frac{(9,03 - 0,9)^2 + (7,47 - 0,9)^2 + \dots + (0 - 0,9)^2}{100-1}} = 2,752.$$

Таким образом, границы доверительного интервала, содержащего полную разность генеральной совокупности, состоящей из 5 000 накладных, вычисляются по формуле (7.10):

$$5\,000 \times 0,90 \pm 1,9842 \times 5\,000 \frac{2,752}{\sqrt{100}} \sqrt{\frac{5\,000-100}{5\,000-1}} = 4\,500 \pm 2\,702,91,$$

$$1\,797,09 \leq \text{Полная разность} \leq 7\,202,91.$$

Следовательно, вероятность того, что полная разность генеральной совокупности накладных не меньше 1 797,09 и не больше 7 202,91, равна 95%. На рис. 7.14 показано, как вычисляется доверительный интервал, содержащий полную разность генеральной совокупности накладных, с помощью программы Microsoft Excel.

	A	B	C	D	E
1	Полная разность между фактическими и зарегистрированными накладными				
2					
3	Data				
4	Population Size	5000			
5	Sample Size	100			
6	Confidence Level	95%			
7					
8	Intermediate Calculations				
9	Sum of Differences	90		Calculations Area	
10	Average Difference in Sample	0.9		For standard deviation of differences:	
11	Total Difference	4500		Number of Differences Not = 0	12
12	Standard Deviation of Differences	2.751707		Number of Differences = 0	88
13	FPC Factor	0.990049		SS for Differences Not = 0	678.3864
14	Standard Error of the Total Diff	1362.306		SS for Differences = 0	71.28
15	Degrees of Freedom	99		Sum of Squares	749.6664
16	t Value	1.984217		Variance of Differences	7.572388
17	Interval Half-Width	2702.913			
18					
19	Confidence Interval				
20	Interval Lower Limit	1797.09			
21	Interval Upper Limit	7202.91			

Рис. 7.14. Вычисление доверительного интервала, содержащего полную разность генеральной совокупности накладных, с помощью программы Microsoft Excel

Процедуры Excel: построение доверительного интервала для полной разности генеральной совокупности ▼

Чтобы построить доверительный интервал для общей суммы генеральной совокупности при неизвестном стандартном отклонении, можно воспользоваться функцией СТЬЮДРАСПОВР или надстройкой PHStat2.

Например, чтобы построить доверительный интервал, показанный на рис. 7.14, необходимо открыть рабочий лист **Данные** в книге **Chapter 7.xls** и выполнить одну из следующих процедур.

Применение Excel в сочетании с надстройкой PHStat2

Чтобы построить доверительный интервал для полной разности генеральной совокупности, следует применить процедуру PHStat⇒Confidence Intervals⇒Estimate for the Total Difference... (PHStat⇒Доверительные интервалы⇒Оценка полной разности...), придерживаясь инструкций, приведенных ниже.

1. Выбрать PHStat⇒Confidence Intervals⇒Estimate for the Total Difference....
2. В диалоговом окне Estimate for the Total Difference (см. иллюстрацию) сделать следующее.
 - 2.1. Ввести в окне редактирования Sample Size (Объем выборки) число 100.
 - 2.2. Ввести в окне редактирования Population Size (Объем генеральной совокупности) число 5000.
 - 2.3. Ввести в окне редактирования Confidence Level (Доверительный уровень) число 95.
 - 2.4. Ввести в окне редактирования Differences Cell Range (Диапазон ячеек, содержащий разности) диапазон A1:A13 и установить флажок First cell contains label (Первая ячейка содержит метку).
 - 2.5. Ввести в окне редактирования Title соответствующий заголовок.
 - 2.6. Щелкнуть на кнопке ОК.

**Применение Excel**

Для того чтобы самостоятельно создать рабочий лист, вычисляющий доверительный интервал для полной разности элементов генеральной совокупности, следуйте инструкциям, приведенным в разделе ЕН.7.7.

 Chapter 7.xls

Данные, на основе которых вычисляется доверительный интервал, содержащий полную разность, как показано на рис. 7.14, хранится в рабочей книге Chapter 7 на листе Рис7.14.

В рассмотренном ранее примере все 12 разностей были положительными, поскольку общая сумма накладных превышала сумму, записанную в базе данных. Разумеется, ошибки могут быть не только положительными, но и отрицательными. Проиллюстрируем это с помощью следующего примера.

ПРИМЕР 7.8. ОЦЕНКА ПОЛНОЙ РАЗНОСТИ

Вернемся к примеру 7.7. Предположим, что в выборке, состоящей из 100 записей, обнаружены 14 накладных, содержащих ошибки. Перечислим их разности, выраженные в долларах.

 DIFFTEST.XLS

75,41	38,97	108,54	-37,18	62,75	118,32	-88,84
127,74	55,42	39,03	29,41	47,99	28,73	84,05

Необходимо построить 95%-ный доверительный интервал, содержащий полную разность генеральной совокупности, состоящей из 1 000 накладных.

РЕШЕНИЕ.

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n} = \frac{690,34}{50} = 13,8068,$$

$$S_D = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}} = \sqrt{\frac{(75,41 - 13,8068)^2 + (38,97 - 13,8068)^2 + \dots + (0 - 13,8068)^2}{50-1}} = 37,427$$

Затем по формуле (7.10) вычисляем границы 95%-ного доверительного интервала, содержащего полную разность.

$$1\,000 \times 13,8068 \pm 1,000 \times 2,0096 \frac{37,427}{\sqrt{50}} \sqrt{\frac{1\,000-50}{1\,000-1}} = 13\,806,8 \pm 10\,372,4.$$

$$3\,434,40 \leq \text{Полная разность} \leq 24\,179,20.$$

Следовательно, вероятность того, что полная разность генеральной совокупности не меньше 3 434,40 и не больше 24 179,20, равна 95%. ■

Односторонняя оценка доли нарушений установленных правил

Организации часто используют механизмы внутреннего контроля, чтобы убедиться в том, что сотрудники следуют правилам, принятым в компании. Например, администрация компании Saxon Home Improvement Company настаивает, чтобы товары не отгружались со склада без заверенной складской расписки. В ходе ежемесячной проверки ревизоры попытались определить долю продукции, отправленной со склада без соответствующих документов. Такое явление мы будем называть *нарушением правил*. Чтобы оценить их количество, ревизоры извлекают из генеральной совокупности накладных выборку и подсчитывают количество нарушений. Затем аудиторы сравнивают результаты с предельно допустимым уровнем нарушений, принятым в компании. Этот уровень представляет собой максимальную долю нарушений, которую компания считает приемлемой. При оценке количества нарушений обычно применяют **односторонний доверительный интервал** (one-sided confidence interval). Иначе говоря, ревизоры вычисляют только верхнюю границу доверительного интервала, содержащего долю нарушений.

ОДНОСТОРОННИЙ ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ, СОДЕРЖАЩИЙ ДОЛЮ ПРИЗНАКА В ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

$$\text{Верхняя граница} = p_s + Z \sqrt{\frac{p_s(1-p_s)}{n}} \sqrt{\frac{N-n}{N-1}}, \quad (7.11)$$

где Z — критическое значение стандартизованного нормального распределения, соответствующее интегральной вероятности, равной $1-\alpha$.

Если предельно допустимый уровень нарушений выше верхней границы этого интервала, ревизоры считают, что деятельность компании не выходит за пределы установленных норм. В противном случае аудиторы фиксируют систематическое нарушение установленных правил. Аудиторы могут потребовать увеличить объем выборки. Предположим, что ежемесячная проверка коснулась 400 накладных, извлеченных из генеральной совокупности, состоящей из 10 000 накладных. Среди проверенных накладных оказалось 20 неправильных. Предельно допустимый уровень нарушений, принятый в компании, равен 6%. Постройте односторонний 95%-ный доверительный интервал.

Итак, $p_s = 20/400 = 0,05$, а $Z = 1,645$. Используя формулу (7.11), получаем:

$$\begin{aligned} \text{Верхняя граница} &= p_s + Z \sqrt{\frac{p_s(1-p_s)}{n}} \sqrt{\frac{N-n}{N-1}} = 0,05 + 1,645 \sqrt{\frac{0,05(1-0,05)}{400}} \sqrt{\frac{10\,000-400}{10\,000-1}} = \\ &= 0,05 + 1,645 \times 0,0109 \times 0,98 = 0,05 + 0,0175 = 0,0675. \end{aligned}$$

Итак, верхняя граница доверительного интервала равна 6,76%. Поскольку предельно допустимый уровень нарушений равен 6%, следует зафиксировать существенное нарушение правил компании и потребовать для проверки более крупную выборку.

ПРИМЕР 7.9. ОЦЕНКА УРОВНЯ НАРУШЕНИЙ

Работники крупной фирмы, производящей электронную бытовую аппаратуру, подписывают около одного миллиона чеков в год. В соответствии с правилами, принятыми в компании, любой чек становится действительным только после того, как будет подписан бухгалтером. Компания считает приемлемыми не более 4% нарушений. Предположим, что в ходе проверки выяснилось, что в выборке, состоящей из 400 чеков, восемь оказались недействительными. Какой вывод должен сделать аудитор?

РЕШЕНИЕ. Следует построить 95%-ный доверительный интервал, содержащий долю нарушений, и сравнить его верхнюю границу с предельно допустимым уровнем ошибок.

В данном случае $p_s = 8/400 = 0,02$, а $Z = 1,645$. Используя формулу (7.11), получаем:

$$\begin{aligned} \text{Верхняя граница} &= p_s + Z \sqrt{\frac{p_s(1-p_s)}{n}} \sqrt{\frac{N-n}{N-1}} = 0,02 + 1,645 \sqrt{\frac{0,02(1-0,02)}{400}} \sqrt{\frac{1\,000\,000-400}{1\,000\,000-1}} = \\ &= 0,02 + 1,645 \times 0,007 \times 0,9998 = 0,02 + 0,0115 = 0,0315. \end{aligned}$$

Итак, верхняя граница доверительного интервала равна 3,15%. Поскольку предельно допустимый уровень нарушений равен 4%, следует признать, что правила компании не нарушены. Иначе говоря, аудитор на 95% уверен, что количество нарушений не превышает 4%. ■

УПРАЖНЕНИЯ К РАЗДЕЛУ 7.5

Изучение основ

- 7.50.** Из генеральной совокупности, состоящей из 500 элементов, извлекается выборка, объем которой равен 25. Выборочное среднее равно 25,7, а выборочное стандартное распределение — 7,8. Постройте 99%-ный доверительный интервал, содержащий общую сумму элементов генеральной совокупности.
- 7.51.** Из генеральной совокупности, состоящей из 10 000 элементов, извлечена выборка, объем которой равен 200. Среди них 10 элементов оказались ошибочными.  TEMERR.XLS.

13,76	42,87	34,65	11,09	14,54
22,87	25,52	9,81	10,03	15,49

Постройте 99%-ный доверительный интервал, содержащий полную разность генеральной совокупности.

- 7.52.** Предположим, что $p_s = 0,04$, $n = 300$, а $N = 5000$. Вычислите верхнюю границу одностороннего доверительного интервала, содержащего долю признака p с заданной вероятностью.
1. 90%.
 2. 95%.
 3. 99%.

Применение понятий

7.53. Магазин канцелярских принадлежностей желает оценить общую стоимость 300 поздравительных открыток, хранящихся на складе. Выборочное среднее, вычисленное по случайной выборке, состоящей из 20 поздравительных открыток, равно 1,67 долл., а выборочное стандартное отклонение — 0,32 долл. Постройте интервал, содержащий общую стоимость поздравительных открыток, доверительный уровень которого равен 95%.

7.54. В крупной компании работают 3 000 сотрудников. Администрация этой компании решила провести опрос и выяснить среднегодовую величину расходов на медицинские услуги. Для этого из генеральной совокупности сотрудников извлечена выборка, объем которой равен 10.

 DENTAL.XLS

110 362 246 85 510 208 173 425 316 179

Постройте 90%-ный доверительный интервал, содержащий общую сумму, затраченную сотрудниками компании на медицинские услуги.

7.55. Подразделение крупной складской сети проводит ежемесячную инвентаризацию товаров. Выяснилось, что на момент проверки на складе хранилось 1 546 предметов. Для проверки из них были случайным образом выбраны 50 предметов и вычислены средняя стоимость $\bar{X} = 252,28$ и стандартное отклонение $S = 93,67$. Постройте 95%-ный доверительный интервал, содержащий общую сумму товаров, хранящихся на складе.

7.56. Клиенты оптовых магазинов одежды часто получают скидки на приобретаемые ими товары. Из генеральной совокупности, состоящей из 4 000 накладных, извлечена случайная выборка, содержащая 150 элементов. Оказалось, что в 13 случаях клиенты не получили скидок, которые им полагались. Объемы скидок, не полученных покупателями, перечислены в таблице.  DISCOUNT.XLS.

6,45 15,32 97,36 230,63 104,18 84,92 132,76
66,12 26,55 129,43 88,32 47,81 89,01

Постройте 99%-ный доверительный интервал, содержащий общую сумму скидок, не полученных клиентами.

7.57. Esone Dresses — небольшая компания, производящая женскую одежду для продажи через сеть специальных магазинов. В инвентарной ведомости этой компании хранятся 1 200 записей, внесенных по принципу FIFO — “первой внесена, первой удалена”. Известно, что около 15% записей являются неверными. В ходе проверки из генеральной совокупности инвентарных записей извлечена выборка, содержащая 120 элементов. Разница между реальной и фиктивной ценой каждого товара приведена в таблице.  FIFO.XLS.

Номер образца	Фактическая стоимость	Фиктивная стоимость
5	261	240
9	87	105
17	201	276
18	121	110
28	315	298

Номер образца	Фактическая стоимость	Фиктивная стоимость
35	411	356
43	249	211
51	216	305
60	21	210
73	140	152
86	129	112
95	340	216
96	341	402
107	135	97
119	228	220

Постройте 95%-ный доверительный интервал, содержащий полную разность между фактической и фиктивной ценами.

7.58. Компания Tom and Brent's Alpline Outfitters проводит ежегодный аудит своей финансовой отчетности. В соответствии с правилами, принятыми в компании, любой чек становится действительным только после того, как будет подписан бухгалтером. Компания считает приемлемыми не более 4% нарушений. В ходе проверки из генеральной совокупности, состоящей из 10 000 записей, были извлечены 300 чеков, среди которых 11 оказались недействительными.

1. Постройте 95%-ный односторонний доверительный интервал, содержащий долю нарушений в финансовой отчетности.
2. Какой вывод должен сделать аудитор?

7.59. Компания Rhonda's Online Fashion Accessories требует, чтобы на каждую партию товара заполнялся гарантийный талон. Фирма допускает не более 5% нарушений. В ходе проверки из генеральной совокупности, состоящей из 5 000 талонов, были извлечены 500 талонов, среди которых 12 оказались недействительными.

1. Постройте 95%-ный односторонний доверительный интервал, содержащий долю нарушений.
2. Какой вывод должен сделать аудитор?

7.6. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ И ЭТИЧЕСКИЕ ПРОБЛЕМЫ



При выборочном исследовании генеральной совокупности и формулировании статистических выводов часто возникают этические проблемы. Основная из них — как согласуются доверительные интервалы и точечные оценки выборочных статистик. Публикация точечных оценок без указания соответствующих доверительных интервалов (как правило, имеющих 95%-ный доверительный уровень) и объема выборки, на основе которых они получены, может породить недоразумения. Это может создать у пользователя впечатление, что точечная оценка — именно то, что ему необходимо, чтобы предсказать свойства всей генеральной совокупности. Таким образом, необходимо понимать, что в любых исследованиях во главу угла должны быть поставлены не точечные, а интервальные оценки. Кроме того, особое внимание следует уделять правильному выбору объемов выборки.

Чаще всего объектами статистических манипуляций становятся результаты социологических опросов населения по тем или иным политическим проблемам. При этом результаты опроса выносятся на первые страницы газет, а ошибку выборочного исследования и методологию статистического анализа печатают где-нибудь в середине. Чтобы доказать обоснованность полученных точечных оценок, необходимо указывать объем выборки, на основе которой они получены, границы доверительного интервала и его уровень значимости.

7.7. ВЫЧИСЛЕНИЕ ОЦЕНОК И ОБЪЕМА ВЫБОРОК, ИЗВЛЕЧЕННЫХ ИЗ КОНЕЧНОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Оценка математического ожидания

В разделе 6.9 поправочный коэффициент для конечной генеральной совокупности (*fpc*) использовался для уменьшения стандартной ошибки в $\sqrt{(N-n)/(N-1)}$. При вычислении доверительных интервалов для оценок параметров генеральной совокупности поправочный коэффициент применяется в ситуациях, когда выборки извлекаются без возвращения. Таким образом, доверительный интервал для математического ожидания, имеющий доверительный уровень, равный $(1-\alpha)\times 100\%$, вычисляется по формуле (7.12).

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ КОНЕЧНОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ (σ НЕИЗВЕСТНА)

$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad (7.12)$$

Чтобы проиллюстрировать применение поправочного коэффициента для конечной генеральной совокупности, вернемся к задаче о вычислении доверительного интервала для средней суммы накладных в компании Saxon Home Improvement Company, рассмотренной в разделе 7.2. Предположим, что за месяц в компании выписываются 5 000 накладных, причем $\bar{X} = 110,27$ долл., $S = 28,95$ долл., $N = 5\,000$, $n = 100$, $\alpha = 0,05$, $t_{99} = 1,9842$. По формуле (7.12) получаем следующие результаты.

$$110,27 \pm 1,9842 \frac{28,95}{\sqrt{100}} \sqrt{\frac{5\,000 - 100}{5\,000 - 1}} = 110,27 \pm 5,744 \times 0,99 = 110,27 \pm 5,69$$

$$104,58 \leq \mu \leq 115,96.$$

Результаты решения этой задачи с помощью программы PHStat представлены на рис. 7.15.

Поскольку в данной задаче выборка представляет собой очень маленькую часть генеральной совокупности, поправочный коэффициент почти не влияет на ширину доверительного интервала. Для того чтобы проверить влияние поправочного коэффициента на ширину доверительного интервала, когда объем выборки превышает 5% генеральной совокупности, рассмотрим следующий пример.

	A	B
1	Средняя сумма накладных	
2		
3	Data	
4	Sample Standard Deviation	28.95
5	Sample Mean	110.27
6	Sample Size	100
7	Confidence Level	95%
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	2.895
11	Degrees of Freedom	99
12	t Value	1.984217306
13	Interval Half Width	5.744309101
14		
15	Confidence Interval	
16	Interval Lower Limit	104.53
17	Interval Upper Limit	116.01
18		
19		
20	Finite Populations	
21	Population Size	5000
22	FPC Factor	0.990049503
23	Interval Half Width	5.687144529
24	Interval Lower Limit	104.58
25	Interval Upper Limit	115.96

Рис. 7.15. Оценка доверительного интервала для средней суммы накладных компании Saxon Home Improvement Company с учетом поправочного коэффициента для конечной генеральной совокупности

ПРИМЕР 7.10. ОЦЕНКА СРЕДНЕГОДОВОВОГО ПОТРЕБЛЕНИЯ ТОПЛИВА

В примере 7.3 из генеральной совокупности была извлечена выборка, состоящая из 30 изоляторов. Допустим, что вся генеральная совокупность состоит из 300 изоляторов. Постройте 95%-ный доверительный интервал для математического ожидания генеральной совокупности.

РЕШЕНИЕ. Поскольку $\bar{X} = 1\,723,4$ галлонов, $S = 89,55$, $n = 30$, $N = 300$ и $t_{29} = 2,0452$ (для доверительного уровня, равного 95%), с учетом поправочного коэффициента для конечной генеральной совокупности получаем следующие результаты.

$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = 1\,723,4 \pm 2,0452 \frac{89,55}{\sqrt{30}} \sqrt{\frac{300-30}{300-1}} = 1\,723,4 \pm 33,44 \times 0,9503 = 1\,723 \pm 31,776.$$

$$1\,691,62 \leq \mu \leq 1\,755,18.$$

Объем выборки в этой задаче равен 10% объема генеральной совокупности, поэтому поправочный коэффициент оказывает небольшое влияние на ширину доверительного интервала. ■

Оценка доли признака

При выборе без возвращения доверительный интервал для доли признака, имеющий доверительный уровень, равный $(1-\alpha) \times 100\%$, вычисляется по формуле (7.13).

**ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ДОЛИ ПРИЗНАКА В КОНЕЧНОЙ
ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ**

$$p_s \pm Z \sqrt{\frac{p_s(1-p_s)}{n}} \sqrt{\frac{N-n}{N-1}}. \quad (7.13)$$

Чтобы проиллюстрировать применение поправочного коэффициента при вычислении доверительного интервала для доли признака в конечной генеральной совокупности, вновь вернемся к задаче о накладных в компании Saxon Home Improvement Company, рассмотренной в разделе 7.3. Исходные данные таковы: $N = 5\,000$, $n = 100$, $p_s = 10/100 = 0,10$, $\alpha = 0,05$, $Z = 1,96$. По формуле (7.13) получаем следующие результаты.

$$\begin{aligned} p_s \pm Z \sqrt{\frac{p_s(1-p_s)}{n}} \sqrt{\frac{N-n}{N-1}} &= 0,10 \pm 1,96 \sqrt{\frac{0,10 \times 0,90}{100}} \sqrt{\frac{5\,000 - 100}{5\,000 - 1}} = \\ &= 0,10 \pm 1,96 \times 0,03 \times 0,99 = 0,10 \pm 0,0582 \\ &0,0418 \leq p \leq 0,1582. \end{aligned}$$

В рассмотренной задаче выборка представляет собой очень маленькую часть генеральной совокупности, поэтому поправочный коэффициент почти не влияет на ширину доверительного интервала.

Определение объема выборки

Поправочный коэффициент можно также применять для определения объема выборки, извлеченной из конечной генеральной совокупности без возвращения. Например, при оценке математического ожидания выборочная ошибка вычисляется по следующей формуле.

$$e = Z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}.$$

При оценке доли признака ошибка выборочного исследования равна

$$e = Z \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}.$$

Чтобы вычислить объем выборки для оценки математического ожидания или доли признака, применяются формулы (7.4) и (7.5):

$$n_0 = \frac{Z^2 \sigma^2}{e^2} \quad \text{и} \quad n_0 = \frac{Z^2 p(1-p)}{e^2},$$

где n_0 — объем выборки без учета поправочного коэффициента для конечной генеральной совокупности. Применение поправочного коэффициента приводит к следующей формуле.

**ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ С УЧЕТОМ ПОПРАВОЧНОГО КОЭФФИЦИЕНТА
ДЛЯ КОНЕЧНОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ**

$$n = \frac{n_0 N}{n_0 + (N-1)}. \quad (7.14)$$

При решении задачи о компании Saxon Home Improvement Company оказалось, что для оценки средней суммы накладных необходима выборка, состоящая из 97 накладных, а для оценки доли ошибочных накладных необходима выборка, состоящая из 100 записей. Используя формулу (7.14) для оценки математического ожидания при $N = 5\,000$, $e = 5$ долл., $S = 25$ долл. и $Z = 1,96$ (для доверительного уровня, равного 95%), получаем следующие результаты.

$$n = \frac{96,04 \times 5\,000}{96,04 + (5\,000 - 1)} = 94,25.$$

Таким образом, $n = 95$.

Применение формулы (7.14) для оценки доли признака при $N = 5\,000$, $e = 0,07$, $p = 0,15$ и $Z = 1,96$ (для доверительного уровня, равного 95%), получаем следующие результаты.

$$n = \frac{99,96 \times 5\,000}{99,96 + (5\,000 - 1)} = 98,02.$$

Таким образом, $n = 99$.

Итак, для того чтобы правильно вычислить обе оценки необходимо выбрать наибольший из двух объемов, т.е. 99. Результаты вычислений, полученных с помощью программы PHStat, представлены на рис. 7.16.

	A	B
1	Объем выборки для оценки среднего	
2		
3	Data	
4	Population Standard Deviation	25
5	Sampling Error	5
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Z Value	-1.95996279
10	Calculated Sample Size	96.036312
11		
12	Result	
13	Sample Size Needed	97
14		
15		
16	Finite Populations	
17	Population Size	5000
18	Sample Size with FPC	94.24601274
19	Sample Size Needed	95

Рис. 7.16. Оценка доверительного интервала для средней суммы накладных компании Saxon Home Improvement Company с учетом поправочного коэффициента для конечной генеральной совокупности

Процедуры Excel: вычисление поправочного коэффициента для конечной генеральной совокупности

Чтобы вычислить поправочный коэффициент для конечной генеральной совокупности и требуемый объем выборки, необходимо добавить новые формулы в рабочие листы, описанные в главе. Надстройка PHStat2 позволяет учесть поправку автоматически.

Применение Excel в сочетании с PHStat2

Установите флажок Finite Population Correction (Поправочный коэффициент для конечной генеральной совокупности) и введите объем генеральной совокупности в окне редактирования Population Size (Объем генеральной совокупности). Эти действия следует выполнить, находясь в диалоговых окнах процедур PHStat, предназначенных для вычисления доверительных интервалов и определения объема выборки.

Применение Excel

Чтобы вычислить поправочный коэффициент для конечных генеральных совокупностей при оценке доверительных интервалов и определении объема выборок, введите дополнительные формулы в рабочие листы, описанные в табл. ЕН.7.1–ЕН.7.4. Изменения, которые необходимо внести в рабочие листы, предназначенные для вычисления доверительных интервалов для математического ожидания при известном стандартном отклонении, математического ожидания при неизвестном стандартном отклонении и доли признака, а также для определения объема выборки при оценке математического ожидания и доли признака соответственно, представлены в табл. 7.3–7.6.

Таблица 7.3. Дополнения к табл. ЕН.7.1 при вычислении доверительного интервала для математического ожидания (σ известна)

	А	В
19	Конечные генеральные совокупности	
20	Объем генеральной совокупности	5000
21	Поправочный коэффициент	=КОРЕНЬ((B20-B6)/(B20-1))
22	Половина ширины интервала	=B12*B21
23	Нижняя доверительная граница	=B5-B22
24	Верхняя доверительная граница	=B5+B22

Таблица 7.4. Дополнения к табл. ЕН.7.2 при вычислении доверительного интервала для математического ожидания (σ неизвестна)

	А	В
20	Конечные генеральные совокупности	
21	Объем генеральной совокупности	5000
22	Поправочный коэффициент	=КОРЕНЬ((B21-B6)/(B21-1))
23	Половина ширины интервала	=B13*B22
24	Нижняя доверительная граница	=B5-B23
25	Верхняя доверительная граница	=B5+B23

Таблица 7.5. Дополнения к табл. ЕН.7.3 при вычислении доверительного интервала для доли признака

	А	В
20	Конечные генеральные совокупности	
21	Объем генеральной совокупности	5000
22	Поправочный коэффициент	=КОРЕНЬ((B20-B4)/(B20-1))
23	Половина ширина интервала	=B12*B21
24	Нижняя доверительная граница	=B9-B22
25	Верхняя доверительная граница	=B9+B22

Таблица 7.6. Дополнения к табл. ЕН.7.4 и ЕН.7.5 при вычислении объема выборки для оценки математического ожидания и доли признака

	А	В
16	Конечные генеральные совокупности	
17	Объем генеральной совокупности	5000
18	Объем выборки с учетом поправочного коэффициента	$=(B10-B17)/(B10+B17-1)$
19	Необходимый объем выборки	$=\text{ОКРУГЛВВЕРХ}(B18;0)$

УПРАЖНЕНИЯ К РАЗДЕЛУ 7.7

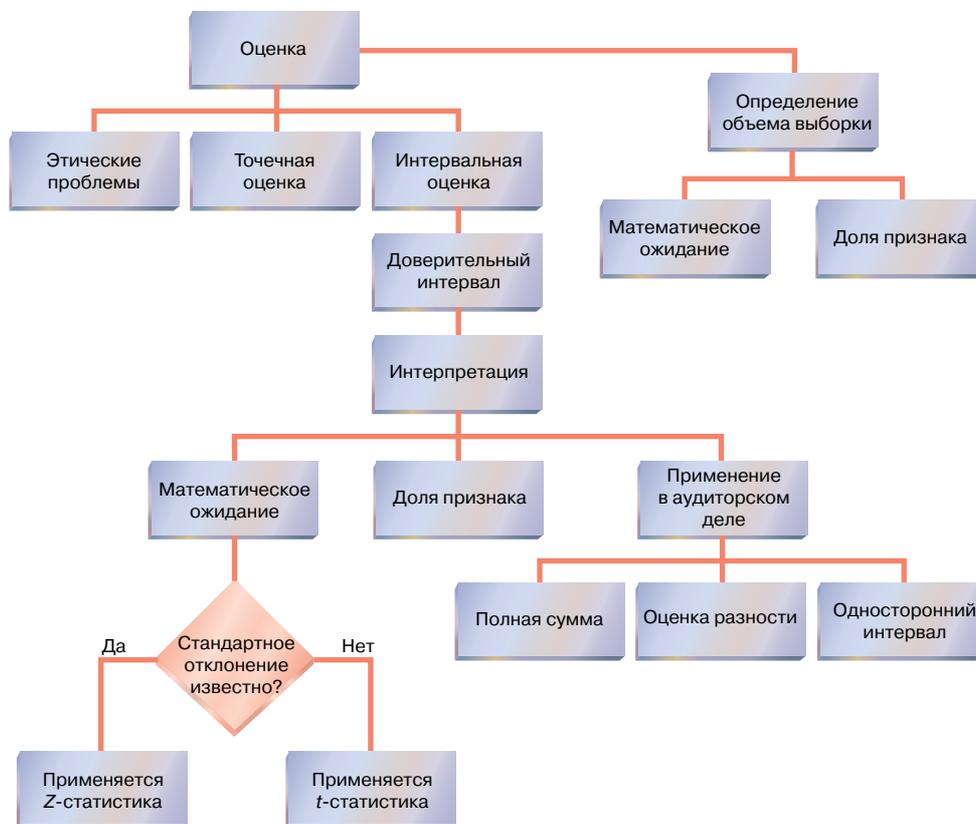
Изучение основ

- 7.60.** Предположим, что $\bar{X} = 75$, $S = 24$, $n = 36$ и $N = 200$, причем выборка получена путем извлечения без возвращения. Постройте 95% -ный доверительный интервал для математического ожидания μ конечной генеральной совокупности.
- 7.61.** Допустим, что объем генеральной совокупности равен 1 000, а стандартное отклонение равно 20. Какой объем выборки необходим, если выбор осуществляется без возвращения, доверительный уровень равен 95%, а выборочная ошибка равна ± 5 .

Применение понятий

- 7.62.** Инспектор отдела технического контроля на фабрике, производящей электрические лампочки, желает оценить среднюю продолжительность работы лампочек из крупной партии. Стандартное отклонение этой величины известно и равно 100 часов. Предположим, что партия состоит из 2 000 электрических лампочек, а выбор выполняется без возвращения.
1. Постройте 95% -ный доверительный интервал для средней продолжительности работы лампочек из указанной партии, если средняя продолжительность работы 50 лампочек, принадлежащих выборке, извлеченной из партии, равно 350 часов.
 2. Определите объем выборки, необходимый для оценки средней продолжительности работы лампочек, если стандартное отклонение равно ± 20 часов, а доверительный уровень равен 95%.
 3. Как изменятся ответы к задачам 1 и 2, если партия состоит из 1 000 лампочек?
- 7.63.** Крупная компания планирует провести статистическое исследование, чтобы определить средний годовой объем средств, затрачиваемых семьями сотрудников на медицинское обслуживание. Администрация желает гарантировать 95% -ный доверительный уровень оценки, допуская выборочную ошибку, равную ± 50 долл. Предварительный опрос показал, что стандартное отклонение равно 400 долл. Какой объем выборки следует установить, если в компании работают 3 000 сотрудников, а извлечение выборки производится без возвращения?
- 7.64.** Управляющий банка, обслуживающего 1 000 жителей маленького городка, желает оценить долю клиентов, имеющих несколько счетов.
1. Постройте 90% -ный доверительный интервал для доли вкладчиков, имеющих несколько счетов, если банк сформировал выборку из 100 клиентов, используя выбор без возвращения и оказалось, что 30 из них имеют несколько счетов.

2. Управляющий банка желает построить 90%-ный доверительный интервал, содержащий долю вкладчиков, имеющих несколько счетов, установив выборочную ошибку равной $\pm 0,05$. Какой объем выборки следует установить, если извлечение выборки производится без возвращения?
 3. Как изменятся ответы к задачам 1 и 2, если банк обслуживает 2 000 вкладчиков?
- 7.65.** Дилер автомобильной компании желает оценить долю клиентов, продолжающих ездить на автомобилях, приобретенных пять лет назад. Записи о продажах указывают, что генеральная совокупность состоит из 4 000 клиентов.
1. Постройте 95%-ный доверительный интервал для доли клиентов, продолжающих ездить на автомобилях, приобретенных пять лет назад, если дилер сформировал выборку из 200 клиентов, используя выбор без возвращения и оказалось, что 82 из них продолжают ездить на старых автомобилях.
 2. Какой объем выборки необходим для оценки истинной доли клиентов, клиентов, продолжающих ездить на автомобилях, приобретенных пять лет назад, если доверительный уровень равен 95%, а выборочная ошибка равна $\pm 0,025$?
 3. Как изменятся ответы к задачам 1 и 2, если количество клиентов равно 6 000?
- 7.66.** Отдел технического контроля крупной компании решил оценить реальный объем безалкогольных напитков, разлитых в двухлитровые бутылки на месте завода. Генеральная совокупность состоит из 2 000 бутылок. Завод проинформировал компанию, что стандартное отклонение объема жидкости, разлитой в двухлитровые бутылки, равно 0,05 л.
1. Постройте 95%-ный доверительный интервал для среднего объема напитка в каждой бутылке, если в выборке, состоящей из 100 двухлитровых бутылок, извлеченной без возвращения, средний объем напитка в бутылке оказался равным 1,99.
 2. Определите объем выборки, необходимый для оценки математического ожидания генеральной совокупности, если доверительный уровень равен 95%, а выборочная ошибка равна $\pm 0,01$ л.
 3. Как изменятся ответы к задачам 1 и 2, если генеральная совокупность состоит из 1 000 бутылок?
- 7.67.** Магазин канцелярских принадлежностей желает оценить среднюю розничную цену поздравительных открыток. Генеральная совокупность состоит из 300 открыток.
1. Постройте 95%-ный доверительный интервал для средней розничной цены поздравительных открыток, если в выборке, состоящей из 20 открыток, извлеченной без возвращения, средняя цена оказалась равной 1,67 долл., а стандартное отклонение равно 0,32 долл.
 2. Как изменится ответ к задаче 1, если генеральная совокупность состоит из 500 открыток?
-



Структурная схема главы 7

РЕЗЮМЕ

Как показано на структурной схеме главы, мы рассмотрели методы построения доверительных интервалов, содержащих параметры генеральной совокупности, а также способ определения объема выборки, необходимого для достижения заданного доверительного уровня. Мы поняли, как на основе небольшой выборки накладных оценить общую стоимость товаров, отгруженных компанией Saxon Home Improvement Company, и долю неверно заполненных накладных. В следующих четырех главах мы рассмотрим методы проверки гипотез о параметрах генеральной совокупности.

ОСНОВНЫЕ ПОНЯТИЯ

Аудит, 480

Доверительный интервал

для доли признака, 466

односторонний, 487

для математического ожидания

при известном стандартном

отклонении, 451

при неизвестном стандартном

отклонении, 458

для общей суммы, 481

для полной разности, 484

односторонний, 487

Доверительный уровень, 451

Критическое значение, 452

Объем выборки	Ошибка выборочного обследования, 471
для оценки доли признака, 474	Распределение
для оценки среднего, 471	стандартизованное нормальное, 456
Оценка	Стьюдента, 456
интервальная, 449	Средняя разность, 484
точечная, 448	Стандартное отклонение разностей, 484
общей суммы, 481	

УПРАЖНЕНИЯ К ГЛАВЕ 7

Проверка знаний

- 7.68. Почему невозможно построить 100%-ный доверительный интервал, позволяющий оценить параметры генеральной совокупности?
- 7.69. Когда для построения доверительного интервала можно применять t -распределение?
- 7.70. Почему при фиксированном объеме выборки n повышение доверительного уровня приводит к расширению доверительного интервала, что в свою очередь снижает точность оценки?
- 7.71. При каких условиях можно использовать односторонний, а не двусторонний интервал?
- 7.72. Когда следует оценивать общую сумму элементов генеральной совокупности, а не ее математическое ожидание?
- 7.73. Чем оценка разностей отличается от оценки математического ожидания?

Задачи к резюме

Задачи 7.74–7.88 можно решать без применения программы Microsoft Excel. Задачи 7.89–7.94 рекомендуется решать с помощью программы Microsoft Excel.

- 7.74. Журнал *Redbook* регулярно проводит социологические опросы с помощью своего Web-сайта (Kravetz, S., “Work Week”, *Wall Street Journal*, April 13, 1999, A1). Посетителям сайта предлагается заполнить электронную анкету. В ходе одного из опросов 665 женщин спросили, что бы они предпочли: четырехдневные выходные или увеличение заработной платы на 20%. Первый вариант выбрали 412 женщин.
1. Назовите генеральную совокупность, из которой извлечена данная выборка респондентов.
 2. Можно ли данную выборку считать случайной?
 3. Можно ли считать корректными статистические выводы, полученные в результате этого опроса?
 4. Как организовать статистически корректный опрос читателей журнала *Redbook*? Какой объем выборки необходим, чтобы оценить долю женщин, предпочитающих четырехдневные выходные, в генеральной совокупности читателей журнала *Redbook* с точностью $\pm 0,02$ и доверительным уровнем 95%?
- 7.75. В настоящее время компании стали затрачивать гораздо больше времени на исследование кандидатур, претендующих за вакантные должности. В результате выяснилось, что многие кандидаты искажают сведения о себе. Исследование, проведенное компанией Automatic Data Processing, выявило неточности в 44% проверенных резюме. К сожалению, в статье не указан объем выборки, на основе которой сделаны эти выводы (Stephanie Armour, “Security Checks Worry Workers”, *USA Today*, June 19, 2002, B1).

1. Допустим, что выборка состояла из 500 резюме. Постройте 95% -ный доверительный интервал для доли резюме, содержащих неточности.
 2. Можно ли на основании решения задачи 1 утверждать, что половина всех кандидатов допускают неточности, излагая сведения о себе?
 3. Допустим, что выборка состояла из 200 резюме. Постройте 95% -ный доверительный интервал для доли резюме, содержащих неточности.
 4. Можно ли на основании решения задачи 2 утверждать, что половина всех кандидатов допускают неточности, излагая сведения о себе?
 5. Опишите влияние, которое оказывает объем выборки на решения задач 1–3.
 6. Изложите возникающие этические проблемы, если объем выборки не указан.
- 7.76.** Журнал, посвященный маркетингу, провел опрос продавцов, допустивших искажения отчетности и совершивших другие неэтичные поступки (D. Haralson and Q. Tian, “Cheating Hearts”, *USA Today*, February 15, 2001, 1A). Выяснилось, что в 58% случаев продавцы искажали отчетность, в 50% случаев — подрабатывали на стороне в рабочее время, в 22% случаев — оформляли посещение стриптиз-баров как посещение ресторанов и в 19% случаев давали клиентам “откат”. Предположим, что в опросе приняло участие 200 менеджеров.
1. Постройте 95% -ный доверительный интервал, содержащий долю менеджеров, допустивших искажение отчетности.
 2. Постройте 95% -ный доверительный интервал, содержащий долю менеджеров, подрабатывавших на стороне в рабочее время.
 3. Постройте 95% -ный доверительный интервал, содержащий долю менеджеров, оформлявших посещение стриптиз-баров как посещение ресторанов.
 4. Постройте 95% -ный доверительный интервал, содержащий долю менеджеров, дававших клиентам “откат”.
 5. Определите объем выборки, необходимый для оценки доли менеджеров, подрабатывавших на стороне в рабочее время, если точность выборочного исследования равна $\pm 0,02$, а доверительный уровень равен 95% .
 6. Можно ли утверждать, что 58% всех менеджеров, искажают отчетность? Аргументируйте свой ответ.
- 7.77.** Компания Starwood Hotels провела опрос 401 руководителей верхнего звена, играющих в гольф (Del Jones, “Many CEOs Bend the Rules (of Golf)”, *USA Today*, June 26, 2002). Результаты опроса приведены ниже.
- 329 менеджеров жульничают, играя в гольф.
 - 329 менеджеров ненавидят тех, кто жульничает, играя в гольф.
 - 289 менеджеров считают, что поведение игроков на поле и в бизнесе одинаково.
 - 80 игроков в гольф готовы поддаться клиенту ради выгодной сделки.
 - 40 прикидываются больными, чтобы сыграть в гольф.
- Постройте 95% -ный доверительный интервал, содержащий долю менеджеров для каждого пункта анкеты. Какие выводы можно сделать на основе этой информации?
- 7.78.** Специалисты по маркетингу, работающие в компании, производящей электронные бытовые приборы, желают изучить привычки телезрителей в небольшом городке. Для этого они создали случайную выборку, состоящую из 40 человек. Каждого респондента попросили сделать подробные записи о телепередачах, которые он просмотрел в течение недели. Среднее время просмотра телепередач оказалось равным $\bar{X} = 15,3$ ч, а стандартное отклонение — $S = 3,8$ ч. Кроме того, 27 телезрителей смотрели вечерние новости хотя бы по выходным.

1. Постройте 95%-ный доверительный интервал, содержащий среднюю продолжительность просмотра телепередач за неделю.
 2. Постройте 95%-ный доверительный интервал, содержащий долю телезрителей, смотрящих вечерние новости хотя бы по выходным.
 3. Предположим, что данный опрос проводится в другом городе. Какую выборку следует создать, чтобы построить 95%-ный доверительный интервал, содержащий среднюю продолжительность просмотра телепередач с точностью ± 2 ч, если стандартное отклонение равно 5 ч?
 4. Предположим, что данный опрос проводится в другом городе. Какую выборку следует создать, чтобы построить 95%-ный доверительный интервал, содержащий долю телезрителей, смотрящих вечерние новости по крайней мере по выходным, с точностью $\pm 0,035$ ч, если стандартное отклонение неизвестно?
 5. Какой объем должна иметь выборка в задачах 3 и 4, если опросы проводятся одновременно?
- 7.79.** Советник по торговле недвижимостью, работающий в органах местного самоуправления, изучает характеристики многоквартирных домов. Для этого он создал случайную выборку, состоящую из 70 домов и вычислил следующие параметры.
- Отапливаемая площадь дома (кв. фут.): $\bar{X} = 1\,759$, $S = 380$.
 - 42 дома подключены к центральной системе отопления.
1. Постройте 95%-ный доверительный интервал, содержащий среднюю отапливаемую площадь.
 2. Постройте 95%-ный доверительный интервал, содержащий долю домов, подключенных к центральной системе отопления.
- 7.80.** Начальник отдела кадров крупной корпорации желает оценить количество рабочих дней, пропущенных за год религиозными сотрудниками. Для этого он создал случайную выборку, состоящую из 25 религиозных сотрудников, и вычислил следующие параметры.
- Отсутствие на рабочем месте: $\bar{X} = 9,7$ дня, $S = 4,0$ дня.
 - 12 религиозных сотрудников отсутствовали на работе более 10 дней.
1. Постройте 95%-ный доверительный интервал, содержащий среднее время отсутствия на рабочем месте религиозных сотрудников.
 2. Постройте 95%-ный доверительный интервал, содержащий долю религиозных сотрудников, отсутствовавших на работе более 10 дней.
Предположим, что начальник отдела кадров пожелал провести опрос в другом подразделении.
 3. Какой объем выборки он должен установить, чтобы оценить среднее время отсутствия на рабочем месте религиозных сотрудников с 95%-ным доверительным уровнем и точностью $\pm 1,5$ дней, если стандартное отклонение равно 4,5 дня?
 4. Какой объем выборки он должен установить, чтобы оценить долю религиозных сотрудников, отсутствовавших на работе более 10 дней с 90%-ным доверительным уровнем и точностью $\pm 0,075$ дня, если стандартное отклонение неизвестно?
 5. Какой объем должна иметь выборка в задачах 3 и 4, если опросы проводятся одновременно?
- 7.81.** Начальник отдела маркетинга универмага Dotty's желает оценить количество денег, потраченных женщинами на приобретение косметики в течение года. Для этого он создал выборку владельцев кредитных карточек и разослал им анкеты. В результате анкетирования должны быть вычислены следующие показатели.

- Количество денег, потраченных женщинами на приобретение косметики в течение года.
 - Доля женщин, приобретающих косметику в универсаме Dotty's.
1. Какой объем выборки следует зафиксировать, чтобы оценить среднее количество денег, потраченных женщинами на приобретение косметики в течение года с 99%-ным доверительным уровнем и точностью ± 5 долл., если стандартное отклонение равно 18 долл. (по данным предыдущих опросов)?
 2. Какой объем выборки следует зафиксировать, чтобы оценить долю женщин, приобретающих косметику в универсаме Dotty's с 90%-ным доверительным уровнем и точностью $\pm 0,045$?
 3. Какой объем должна иметь выборка в задачах 2 и 3, если опросы проводятся одновременно?
- 7.82.** Менеджер книжного магазина, расположенного в районе студенческого городка, желает оценить предпочтения своих клиентов. Его интересуют два показателя: объем денег, потраченных отдельным клиентом на покупки, и популярность видеокассет с учебными материалами. В частности, менеджера интересуют, покупают ли студенты видеокассеты с лекциями по конкретным предметам, например по статистике, бухучету и т.п., а также с материалами, предназначенными для подготовки к выпускным экзаменам, например, GMAT, GRE или LSAT. Для этого он создал случайную выборку, состоящую из 70 человек, и получил следующую информацию.
- Объем потраченных денег: $\bar{X} = 28,52$ долл., $S = 11,39$ долл.
 - Среди опрошенных клиентов оказалось 28 покупателей, предпочитающих кассеты с учебными материалами.
1. Постройте 95%-ный доверительный интервал, содержащий среднее количество денег, потраченных клиентами.
 2. Постройте 90%-ный доверительный интервал, содержащий долю клиентов, предпочитающих видеокассеты с учебными материалами.
Предположим, что менеджер решил провести опрос в магазине, расположенном в другом студенческом городке.
 3. Какой объем выборки он должен установить, чтобы оценить среднее количество потраченных клиентами денег с 95%-ным доверительным уровнем и точностью ± 2 долл., если стандартное отклонение равно 10 долл.?
 4. Какой объем выборки он должен установить, чтобы оценить среднее количество потраченных клиентами денег с 90%-ным доверительным уровнем и точностью $\pm 0,04$ долл., если стандартное отклонение равно 10 долл.?
 5. Какой объем должна иметь выборка в задачах 3 и 4, если опросы проводятся одновременно?
- 7.83.** Менеджер зоомагазина желает оценить предпочтения своих клиентов. Его интересуют два показателя: объем денег, потраченных отдельным клиентом на покупки, и количество животных, принадлежащих ему: одна собака, один кот или несколько собак и/или кошек. Для этого он создал случайную выборку, состоящую из 70 человек, и получил следующую информацию.
- Объем потраченных денег: $\bar{X} = 21,34$ долл., $S = 9,22$ долл.
 - 37 клиентов имеют только собаку.
 - 26 клиентов имеют только кошку.
 - 7 клиентов имеют несколько собак и/или кошек.
1. Постройте 95%-ный доверительный интервал, содержащий среднее количество денег, потраченных клиентами.

2. Постройте 90%-ный доверительный интервал, содержащий долю клиентов, имеющих только одного кота.

Предположим, что менеджер решил провести опрос в другом зоомагазине.

3. Какой объем выборки он должен установить, чтобы оценить среднее количество потраченных клиентами денег с 95%-ным доверительным уровнем и точностью $\pm 1,5$ долл., если стандартное отклонение равно 10 долл.?
4. Какой объем выборки он должен установить, чтобы оценить долю клиентов, имеющих только одного кота, с 90%-ным доверительным уровнем и точностью $\pm 0,045$ долл.?
5. Какой объем должна иметь выборка в задачах 3 и 4, если опросы проводятся одновременно?

- 7.84.** Владелец ресторана европейской кухни желает знать вкусы своих клиентов. В частности, его интересуют два показателя: объем денег, потраченных отдельным клиентом на покупки, и заказывают ли они десерт. Для этого он создал случайную выборку, состоящую из 60 человек, и получил следующую информацию.

- Объем потраченных денег: $\bar{X} = 38,54$ долл., $S = 7,26$ долл.
- 18 клиентов заказали десерт.

1. Постройте 95%-ный доверительный интервал, содержащий среднее количество денег, затрачиваемых клиентами ресторана.
2. Постройте 90%-ный доверительный интервал, содержащий долю клиентов, заказывающих десерт.

Предположим, что конкурент решил провести опрос в своем ресторане (независимо от результатов первого опроса).

3. Какой объем выборки он должен установить, чтобы оценить среднее количество затрачиваемых клиентами денег с 95%-ным доверительным уровнем и точностью $\pm 1,5$ долл., если стандартное отклонение равно 8 долл.?
4. Какой объем выборки он должен установить, чтобы оценить долю клиентов, заказывающих десерт, с 90%-ным доверительным уровнем и точностью $\pm 0,04$ долл.?
5. Какой объем должна иметь выборка в задачах 3 и 4, если опросы проводятся одновременно?

- 7.85.** Представителя крупной сети магазинов, торгующих промышленным оборудованием, заинтересовали характеристики новой продукции под названием Ice Melt. Производители утверждают, что их порошок позволяет растапливать лед даже при температуре ниже 0° по Фаренгейту (-17° по Цельсию. — *Прим. ред.*). Сеть магазинов приобрела большую партию 5-фунтовых упаковок с порошком. Менеджер желает построить 95%-ный доверительный интервал, содержащий долю продукции Ice Melt, соответствующей рекламным обещаниям, с точностью $\pm 0,05$.

1. Сколько упаковок порошка следует проверить?

Какие предположения о доле признака необходимо принять? (Такая проверка называется *разрушающей*, иначе говоря, после тестирования продукция выходит из строя и не пригодна к продаже.)

Предположим, что для тестирования отобраны 50 упаковок, причем 42 из них оказались пригодными к употреблению.

2. Постройте 95%-ный доверительный интервал, содержащий долю качественной продукции.
3. Стоит ли покупать новый порошок?

7.86. В компании проводится аудиторская проверка. Предположим, что для анализа из генеральной совокупности, состоящей из 1 000 документов, отобраны 50. Оказалось, что семь из них не соответствуют правилам, принятым в компании.

1. Постройте 90%-ный односторонний доверительный интервал, содержащий долю документов, не соответствующих правилам.
2. Допустим, что приемлемый уровень ошибок равен 0,15. Какой вывод должен сделать аудитор?

7.87. Аудитору правительственного агентства поручено оценить, правильно ли осуществляется компенсация оплаты визитов к врачу по программе Medicare. Аудиторской проверке подвергаются все компенсации, выплаченные в определенном районе в течение месяца. Генеральная совокупность состоит из 25 056 визитов к врачу. Аудитор желает оценить общий объем компенсаций, выплаченных в течение месяца. Кроме того, он хочет построить 95%-ный доверительный интервал, содержащий средний объем компенсации с точностью ± 5 долл. Основываясь на прошлом опыте, аудитор считает, что стандартное отклонение равно 30 долл.

1. Какой объем должна иметь выборка?

Предположим, что аудит проводится на основе выборки, определенной в задаче 1. Обнаружилось 12 случаев неправильно оплаченной компенсации, причем $\bar{X} = 93,70$ долл., $S = 34,55$ долл. Разности между правильными и неправильными суммами компенсации приведены в таблице.  MEDICARE.XLS.

17 25 14 -10 20 40 35 30 28 22 15 5

2. Постройте 90%-ный односторонний доверительный интервал, содержащий долю неправильных компенсаций.
3. Постройте 95%-ный доверительный интервал, содержащий среднюю сумму компенсации.
4. Постройте 95%-ный доверительный интервал, содержащий общую сумму компенсации.
5. Постройте 95%-ный доверительный интервал, содержащий полную разность между правильными и неправильными компенсациями.

7.88. Крупный склад компьютерной техники проводит ежемесячную инвентаризацию. Аудитор желает оценить среднюю стоимость компьютеров, хранящихся на складе. Для этого он хочет построить 99%-ный доверительный интервал, содержащий среднюю стоимость компьютеров с точностью ± 200 долл. Основываясь на прошлом опыте, аудитор считает, что стандартное отклонение равно 400 долл.

1. Какой объем должна иметь выборка?

Предположим, что аудит проводится на основе выборки, определенной в задаче 1. Оказалось, что $\bar{X} = 3\,054,13$ долл., $S = 384,62$ долл.

2. Постройте 99%-ный доверительный интервал, содержащий общую стоимость 258 компьютеров, хранящихся на складе.

7.89. Одним из показателей качества процесса упаковки чая является вес отдельного пакетика. Если пакетик чая неполон, возникают две проблемы. Во-первых, потребитель чая может не получить требуемой крепости заварки. Во-вторых, компания может привлечь к ответственности за нарушение правил маркировки. В данном примере на упаковке указывается номинальный средний вес чая в пакетике: 5,5 г. Если реальный средний вес чая в пакетике превышает указанное значение, компания несет дополнительные убытки. Точно засыпать в пакетик

5,5 г невозможно, поскольку температура и влажность воздуха на чайной фабрике постоянно изменяются, а это влияет на плотность чая. Кроме того, скорость работы упаковочной машины чрезвычайно высока (170 пакетиков в минуту). В следующей таблице приведен вес в граммах 50 пакетиков чая, заполненных в течение часа конкретной упаковочной машиной. ● TEABAGS.XLS.

5,65	5,44	5,42	5,40	5,53	5,34	5,54	5,45	5,52	5,41
5,57	5,40	5,53	5,54	5,55	5,62	5,56	5,46	5,44	5,51
5,47	5,40	5,47	5,61	5,53	5,32	5,67	5,29	5,49	5,55
5,77	5,57	5,42	5,58	5,58	5,50	5,32	5,50	5,53	5,58
5,61	5,45	5,44	5,25	5,56	5,63	5,50	5,57	5,67	5,36

1. Постройте 99%-ный доверительный интервал, содержащий средний вес пакетиков с чаем.
2. Соответствует ли средний вес пакета требованиям стандарта?

7.90. Промышленная компания на Среднем Западе производит стальные корпуса для электротехнического оборудования. Основным компонентом корпуса является прямоугольный профиль, который создается из 14-дюймового рулона стальной полосы с помощью 250-тонного пресса. Основным параметром корпуса является расстояние между боковыми сторонами профиля, допускающее установку электротехнического оборудования. В таблице приведены данные о 49 профилях. ● TROUGH.XLS.

8,312	8,343	8,317	8,383	8,348	8,410	8,351	8,373	8,481	8,422
8,476	8,382	8,484	8,403	8,414	8,419	8,385	8,465	8,498	8,447
8,436	8,413	8,489	8,414	8,481	8,415	8,479	8,429	8,458	8,462
8,460	8,444	8,429	8,460	8,412	8,420	8,410	8,405	8,323	8,420
8,396	8,447	8,405	8,439	8,411	8,427	8,420	8,498	8,409	

1. Постройте 95%-ный доверительный интервал, содержащий среднее расстояние между боковыми сторонами профиля.
2. Дайте интерпретацию этого интервала.

7.91. Исследования, проведенные компаниями, производящими рубероидную кровельную плитку в Бостоне и Вермонте, показали, что основным фактором, влияющим на оценку качества продукции, является ее вес. На последнем этапе плитка пакуется, а затем размещается на деревянных стеллажах (как правило, на поддоне помещается 16 плиток). После заполнения стеллажа регистрируется его вес. Файл ● PALLET.XLS содержит данные о весе (в фунтах) 368 стеллажей, заполненных плитками, произведенными в Бостоне, и 330 стеллажей, загруженных плитками, сделанными в Вермонте.

1. Постройте 95%-ный доверительный интервал, содержащий средний вес плиток, произведенных на заводе в Бостоне.
2. Постройте 95%-ный доверительный интервал, содержащий средний вес плиток, произведенных на заводе в Вермонте.
3. Выполняются ли предположения, необходимые для решения задач 1 и 2?
4. Какие выводы о среднем весе плиток, произведенных в Бостоне и Вермонте, можно сделать на основании решения задач 1–3?

7.92. Производитель рубероидной кровельной плитки на заводах в Бостоне и Вермонте предоставляет своим клиентам 20-летнюю гарантию. Для того чтобы убедиться в том, что плитки прослужат указанный срок, на заводах проводят ускоренное испытание на долговечность. В ходе этого эксперимента плитка на протяжении несколь-

ких минут подвергается интенсивному воздействию, эквивалентному воздействию, которому плитка подвергалась бы в обычных условиях в течение 20 лет. В частности, плитку несколько минут очень энергично скребут щетками, а затем взвешивают гранулы, которые отскакивают от плиток (в граммах). Чем меньше гранул образуется в ходе эксперимента, тем долговечнее плитка. Для того чтобы прослужить весь гарантийный срок, плитка не должна потерять больше 0,8 г. В файле  GRANULE.XLS содержатся данные о выборке, состоящей из 170 измерений, проведенных на заводе в Бостоне, и 140 измерениях, осуществленных на заводе в Вермонте.

1. Постройте 95%-ный доверительный интервал, содержащий средний вес гранул, потерянных кровельными плитками, произведенными на заводе в Бостоне.
2. Постройте 95%-ный доверительный интервал, содержащий средний вес гранул, потерянных кровельными плитками, произведенными на заводе в Вермонте.
3. Выполняются ли предположения, необходимые для решения задач 1 и 2?
4. Какие выводы о среднем весе гранул, потерянных кровельными плитками, произведенными в Бостоне и Вермонте, можно сделать на основании решения задач 1–3?

7.93. Компания Zagat публикует рейтинги ресторанов, расположенных в разных городах США. В файле  RESTRATE.XLS содержатся оценки качества пищи, оформления блюд, уровня обслуживания и стоимость обеда для одного человека в 50 ресторанах Нью-Йорк Сити и 50 ресторанах Лонг-Айленда.

Источник: цитируется по изданиям Zagat Survey “2002 New York City Restraunts” и Zagat Survey “2000 long Island Restraunts”.

Выполните следующие задания для ресторанов Нью-Йорк Сити и Лонг-Айленда отдельно.

1. Постройте 95%-ные доверительные интервалы, содержащие оценки качества пищи, оформления блюд, уровня обслуживания и стоимость обеда для одного человека.
2. Какие выводы можно сделать о ресторанах Нью-Йорк Сити и Лонг-Айленда на основе результатов, полученных при решении задачи 1?

Отчеты

7.94. Напишите отчет, содержащий результаты решения задачи 7.82 (расстояния между боковыми сторонами стального профиля). Вычислите статистические показатели с помощью программы Microsoft Excel и вставьте их в отчет, созданный с помощью пакета Microsoft Office, руководствуясь инструкциями из приложения 3.

Применение Интернет

7.95. Зайдите на сайт www.prenhall.com/levine. Выберите ссылку Chapter 7 и щелкните на ссылке Internet exercises.



ГРУППОВОЙ ПРОЕКТ

ТР.7.1. Вернитесь к заданию ТР.2.1. Получите интервальные оценки статистических показателей, характеризующих генеральную совокупность взаимных фондов с очень низким, низким, средним, высоким и очень высоким уровнями риска. Изложите полученные оценки в письменном отчете и доложите на семинаре.

 MUTUAL FUNDS.XLS.

РАЗБОР КОНКРЕТНОЙ СИТУАЦИИ — ГАЗЕТА *SPRINGVILLE HERALD*

Отдел маркетинга газеты *Springville Herald* ищет способы увеличить общее количество подписчиков газеты и удержать подписчиков, согласившихся участвовать в эксперименте. Лорен Альфонсо (Lauren Alfonso), менеджер по продажам, предложила провести опрос, чтобы выяснить привычки читателей, предпочитающих не подписывать газету, а покупать ее в киосках. После продолжительной дискуссии было решено включить в анкету следующие вопросы.

Вы или члены вашей семьи когда-нибудь покупали газету *Springville Herald*?

1. Да.
2. Нет (опрос прекращается).

Выписываете ли вы газету *Springville Herald*?

1. Да.
2. Нет (переходим к вопросу 4).

Когда вы получаете газету *Springville Herald*?

1. С понедельника по субботу.
2. Только в воскресенье.
3. Каждый день (переходим к вопросу 9).

Как часто вы получаете газету *Springville Herald* с понедельника по субботу?

1. Каждый день.
2. Почти каждый день.
3. Редко.

Часто ли вы покупаете газету *Springville Herald* по воскресеньям?

1. Каждое воскресенье.
2. 2–3 раза в месяц.
3. Раз в месяц.

Где вы предпочитаете покупать газету *Springville Herald*?

1. В бакалейных лавках.
2. В кондитерских магазинах.
3. В торговом автомате.
4. В супермаркете.
5. В другом месте.

Согласились бы вы подписаться на газету *Springville Herald* с испытательным сроком по сниженной цене?

1. Да.
2. Нет (переходим к вопросу 9).

В данный момент экземпляр газеты *Springville Herald* стоит 50 центов — с понедельника по субботу и 1 доллар 50 центов — в воскресенье. Суммарная стоимость газеты за

неделю — 4 доллара 50 центов. Какую цену вы готовы заплатить, чтобы получать газету в течение 90 дней?

Читаете ли вы еще какую-либо ежедневную газету, кроме *Springville Herald*?

1. Да.
2. Нет.

В качестве поощрения постоянные подписчики (выписывающие газету в течение 6 месяцев на сумму более 100 долларов) получают дисконтную карточку для посещения некоторых ресторанов в г. Спрингвилль. Хотите ли вы получить такую карточку?

1. Да.
2. Нет.

Для телефонного опроса была создана случайная выборка, состоящая из 500 местных жителей. Телефонный номер генерировался с помощью случайных цифр. При этом первые три цифры фиксировались, а последние четыре — выбирались случайным образом. В опросе принимали участие только телефонные номера, зарегистрированные в г. Спрингвилль.

Из 500 опрошенных жителей 94 респондента отказались принимать участие в анкетировании, не стали отвечать на повторный звонок, имели несуществующий номер или снимали квартиру. Результаты представлены в таблице.

Семьи, покупающие газету <i>Springville Herald</i>	Частота
Да	352
Нет	54

Семьи, выписывающие газету <i>Springville Herald</i>	Частота
Да	136
Нет	216

Разновидность подписки	Частота
С понедельника по субботу	18
Только по воскресеньям	25
Каждый день	93

Поведение покупателей, не являющихся подписчиками (понедельник–суббота)	Частота
Покупают каждый день	78
Покупают почти каждый день	95
Покупают редко	43

Поведение покупателей, не являющихся подписчиками (воскресенье)	Частота
Покупают каждое воскресенье	138
Покупают 2–3 раза в месяц	54
Покупают 1 раз в месяц	24

Место покупки	Частота
Бакалейная лавка	74
Кондитерский магазин	95
Торговый автомат	21
Супермаркет	13
Другое	13

Согласны подписаться на испытательный период	Частота
Да	46
Нет	170

Суммы, которые респонденты согласны платить (в неделю) за подписку в течение 90-дневного испытательного срока.

 SH7.XLS

4,15	3,60	4,10	3,60	3,60	3,60	4,40
3,15	4,00	3,75	4,00	3,25	3,75	3,30
3,75	3,65	4,00	4,10	3,90	3,50	3,75
3,00	3,40	4,00	3,80	3,50	4,10	4,25
3,50	3,90	3,95	4,30	4,20	3,50	3,75
3,30	3,85	3,20	4,40	3,80	3,40	3,50
2,85	3,75	3,80	3,90			

Читают другие ежедневные газеты	Частота
Да	138
Нет	214

Хотели бы получить дисконтные карточки для постоянных подписчиков	Частота
Да	66
Нет	286

УПРАЖНЕНИЯ

Некоторые сотрудники отдела маркетинга заинтересовались методом выбора телефонных номеров респондентов с помощью случайных чисел, использованным для опроса. Подготовьте отчет, в котором исследованы следующие вопросы.

1. Преимущества и недостатки метода случайных чисел.
2. Опишите возможные альтернативы, укажите их преимущества и недостатки.

Проанализируйте результаты опроса жителей г. Спрингвилль. Напишите отчет, содержащий выводы маркетингового исследования.

ПРИМЕНЕНИЕ WEB

Примените свои знания о доверительных интервалах и оцените данные о деятельности компании OnCampus!, упомянутые в главе 6.

Помимо всего прочего, компания OnCampus! предоставляет посетителям своего Web-сайта возможность воспользоваться системой электронной торговли OnCampus! Life-Styles. Для обеспечения электронной системы обработки платежей компания OnCampus! заключила соглашения со следующими фирмами.

- PayAFriend (PAF): система электронной обработки платежей, в которой безопасность обеспечивается предварительной регистрацией всех клиентов и компаний, без применения кредитных карточек.
- Continental Banking Company (Cobanco): провайдер услуг по электронной обработке платежей, позволяющих клиентам компании OnCampus! оплачивать товары общепризнанными кредитными карточками, выпускаемыми финансовыми организациями.

Чтобы снизить стоимость системы, руководство компании OnCampus! решило отказаться от услуг одного из партнеров, обеспечивающих обработку электронных платежей. Однако Вирджиния Даффи (Virginia Duffy) из отдела продаж заподозрила, что две формы оплаты пользуются у клиентов неодинаковой популярностью, что влияет на их покупательскую активность. Чтобы проверить свои подозрения, она решила вычислить следующие статистики.

1. Доля клиентов, пользующихся услугами компании PAF, и доля клиентов, использующих кредитные карточки.
2. Средний объем продаж при оплате через компанию PAF и средний объем продаж при оплате с помощью кредитных карточек.

Помогите мисс Даффи провести анализ на основе случайной выборки, состоящей из 50 транзакций. Собранные данные хранятся в файле по адресу www.prenhall.com/Springville/OnCampus_PymtSample.htm. Подведите итоги своих исследований и ответьте, подтверждается ли гипотеза мисс Даффи. Достаточно ли велика выборка, созданная мисс Даффи для проведения корректного анализа, если ошибка выборочного исследования не превышает 3 долл.?

СПРАВОЧНИК ПО EXCEL. ГЛАВА 7

ЕН.7.1. Вычисление доверительного интервала для математического ожидания при известном стандартном отклонении σ

Чтобы вычислить доверительный интервал для математического ожидания при известном стандартном отклонении σ , создадим рабочий лист, использующий функцию ДОВЕРИТ. Вызов этой функции выглядит следующим образом.

ДОВЕРИТ (1–*доверительный уровень*; *стандартное отклонение*; *объем выборки*).

Шаблон этого рабочего листа, соответствующий примеру 7.1, приведен в табл. ЕН.7.1. Для вычисления нижней и верхней доверительной границы ширина доверительного интервала, возвращаемая функцией ДОВЕРИТ, делится пополам и прибавляется к выборочному среднему. На листе также продемонстрированы вычисления стандартной ошибки и величины Z .

Таблица ЕН.7.1. Шаблон рабочего листа Доверительный_интервал

	А	В
1	Вычисление доверительного интервала, содержащего среднюю длину листа	
2		
3	Данные	
4	Выборочное стандартное отклонение	0,02
5	Выборочное среднее	10,998
6	Объем выборки	100
7	Доверительный уровень	0,95
8		
9	Промежуточные вычисления	
10	Стандартная ошибка среднего	=В4/КОРЕНЬ(В6)
11	Величина Z	=НОРМСТОБР((1-В7)/2)
12	Половина доверительного интервала	=ДОВЕРИТ(1-В7;В4;В6)
13		
14	Доверительный интервал	
15	Нижняя доверительная граница	=В5-В12
16	Верхняя доверительная граница	=В5+В12

При реализации этого шаблона ячейку В7 следует отформатировать так, чтобы величина 0,95 была представлена как 95% (см. врезку ЕР.5). В этом случае величина 0,95 будет представлена как 95%. Если при решении аналогичной задачи выборочное среднее не известно и подлежит вычислению, необходимо заменить формулу в ячейке В5 формулой =СРЗНАЧ (*диапазон*).

ЕН.7.2. Вычисление доверительного интервала для математического ожидания при неизвестном стандартном отклонении σ

Чтобы вычислить доверительный интервал для математического ожидания при неизвестном стандартном отклонении σ , создадим рабочий лист, использующий функцию СТЬЮДРАСПОБР. Вызов этой функции выглядит следующим образом.

СТЬЮДРАСПОБР (1 – доверительный уровень; степени_свободы) .

Шаблон этого рабочего листа, соответствующий рис. 7.6, приведен в табл. ЕН.7.2. Для вычисления половины доверительного интервала, содержащего математическое ожидание, t -значение распределения Стьюдента, возвращаемое функцией СТЬЮДРАСПОБР, умножается на стандартную ошибку среднего и делится пополам.

Таблица ЕН.7.2. Шаблон рабочего листа Доверительный_интервал

	А	В
1	Вычисление доверительного интервала, содержащего среднюю сумму накладных	
2		
3	Данные	
4	Выборочное стандартное отклонение	28,95
5	Выборочное среднее	110,27
6	Объем выборки	100
7	Доверительный уровень	0,95
8		
9	Промежуточные вычисления	
10	Стандартная ошибка среднего	=B4/КОРЕНЬ(B6)
11	Степени свободы	=B6-1
12	t -значение	=СТЬЮДРАСПОБР(1-B7;B11)
13	Половина доверительного интервала	=B12*B10
14		
15	Доверительный интервал	
16	Нижняя доверительная граница	=B5-B13
17	Верхняя доверительная граница	=B5+B13

При реализации этого шаблона ячейку В7 следует отформатировать так, чтобы величина 0,95 была представлена как 95% (см. врезку ЕР.5). Если при решении аналогичной задачи выборочное среднее не известно и подлежит вычислению, необходимо заменить формулу в ячейке В5 формулой, использующей функцию =СРЗНАЧ (диапазон) .

ЕН.7.3. Вычисление доверительного интервала для доли признака в генеральной совокупности

Создадим рабочий лист, использующий для вычисления доверительного интервала, содержащего долю признака в генеральной совокупности, функцию НОРМСТОБР, вызов которой имеет вид НОРМСТОБР (вероятность) , где аргумент *вероятность* представляет собой площадь фигуры, ограниченной кривой стандартизованного нормального распределения и лежащей левее числа X .

Шаблон рабочего листа, соответствующий рис. 7.10, приведен в табл. ЕН.7.3. Для вычисления половины доверительного интервала, содержащего долю признака, значение Z , возвращаемое функцией НОРМСТОБР, умножается на стандартную ошибку среднего и делится пополам. Затем эта половина вычитается и добавляется к выборочному среднему. Чтобы половина доверительного интервала всегда была положительной, в ячейке В12 использована функция ABS.

Таблица ЕН.7.3. Шаблон рабочего листа Доверительный_интервал

	А	В
1	Вычисление доверительного интервала, содержащего долю признака	
2		
3	Данные	
4	Объем выборки	100
5	Количество успехов	10
6	Доверительный уровень	0,95
7		
8	Промежуточные вычисления	
9	Выборочная доля признака	=B5/B4
10	Величина Z	=НОРМСТОБР((1-B6)/2)
11	Стандартная ошибка доли признака	=КОРЕНЬ(B9*(1-B9)/B4)
12	Половина доверительного интервала	=ABS(B10*B11)
13		
14	Доверительный интервал	
15	Нижняя доверительная граница	=B9-B12
16	Верхняя доверительная граница	=B9+B12

При реализации этого шаблона ячейку В6 следует отформатировать так, чтобы величина 0,95 была представлена как 95% (см. врезку ЕР.5).

ЕН.7.4. Определение объема выборки для математического ожидания генеральной совокупности

Создадим рабочий лист, использующий функцию НОРМСТОБР для определения объема выборки, необходимой для вычисления доверительного интервала, содержащего математическое ожидание генеральной совокупности. Вызов этой функции имеет вид НОРМСТОБР (*вероятность*), где аргумент *вероятность* представляет собой площадь фигуры, ограниченной кривой стандартизованного нормального распределения и лежащей левее числа X .

Шаблон рабочего листа, соответствующий рис. 7.11, приведен в табл. ЕН.7.4. Для вычисления объема выборки, необходимой для вычисления доверительного интервала, содержащего среднюю сумму накладных, используется значение Z , возвращаемое функцией НОРМСТОБР. Объем выборки округляется с помощью функции ОКРУГЛВВЕРХ.

Таблица ЕН.7.4. Шаблон рабочего листа Объем_выборки

	А	В
1	Вычисление объема выборки, необходимой для оценки средней суммы накладных	
2		
3	Данные	
4	Стандартное отклонение	25
5	Выборочная ошибка	5
6	Доверительный уровень	0,95
7		
8	Промежуточные вычисления	
9	Величина Z	=НОРМСТОБР((1-В6)/2)
10	Вычисленный объем выборки	=((В9*В4)/В5)^2
11		
12	Результаты	
13	Необходимый объем выборки	=ОКРУГЛВВЕРХ(В10;0)

При реализации этого шаблона ячейку В6 следует отформатировать так, чтобы величина 0,95 была представлена как 95% (см. врезку ЕР.5).

ЕН.7.5. Определение объема выборки для оценки доли признака в генеральной совокупности

Создадим рабочий лист, использующий функцию НОРМСТОБР для определения объема выборки, которая необходима для вычисления доверительного интервала, содержащего долю признака в генеральной совокупности. Вызов этой функции имеет вид НОРМСТОБР (*вероятность*), где аргумент *вероятность* представляет собой площадь фигуры, ограниченной кривой стандартизованного нормального распределения и лежащей левее числа X.

Шаблон рабочего листа, соответствующий рис. 7.12, приведен в табл. ЕН.7.5. Для определения объема выборки, необходимой для вычисления доверительного интервала, содержащего долю ошибочных накладных, используется значение Z, возвращаемое функцией НОРМСТОБР. Объем выборки округляется с помощью функции ОКРУГЛВВЕРХ.

Таблица ЕН.7.5. Шаблон рабочего листа Объем_выборки

	А	В
1	Вычисление объема выборки, необходимой для оценки доли ошибочных накладных	
2		
3	Данные	
4	Оценка истинной доли признака	0,15
5	Выборочная ошибка	0,07
6	Доверительный уровень	0,95
7		

Окончание табл. ЕН.7.5

	А	В
8	Промежуточные вычисления	
9	Величина Z	=НОРМСТОБР((1-В6)/12)
10	Вычисленный объем выборки	=(В9^2*В4*(1-В4)/В5)^2
11		
12	Результаты	
13	Необходимый объем выборки	=ОКРУГЛВВЕРХ(В10;0)

При реализации этого шаблона ячейку В6 следует отформатировать так, чтобы величина 0,95 была представлена как 95% (см. врезку ЕР.5).

ЕН.7.6. Вычисление доверительного интервала, содержащего общую сумму элементов генеральной совокупности

Создадим рабочий лист, использующий для вычисления доверительного интервала, который содержит общую сумму элементов генеральной совокупности, функцию СТЬЮДРАСПОБР. Вызов этой функции выглядит следующим образом.

СТЬЮДРАСПОБР (1-*доверительный уровень; степени свободы*).

Шаблон рабочего листа, соответствующий рис. 7.13, приведен в табл. ЕН.7.6. Для вычисления половины доверительного интервала, содержащего общую сумму элементов генеральной совокупности, *t*-значение распределения Стьюдента, которое возвращается функцией СТЬЮДРАСПОБР, умножается на стандартную ошибку среднего. Затем полученная величина вычитается из выборочного среднего и добавляется к нему.

Таблица ЕН.7.6. Шаблон рабочего листа Доверительный_интервал

	А	В
1	Вычисление доверительного интервала, содержащего общую сумму накладных	
2		
3	Данные	
4	Объем генеральной совокупности	5000
5	Выборочное среднее	110,27
6	Объем выборки	100
7	Выборочное стандартное отклонение	28,95
8	Доверительный уровень	0,95
9		
10	Промежуточные вычисления	
11	Сумма элементов генеральной совокупности	=В4*В5
12	Множитель	=КОРЕНЬ((В4-В6)/(В4-1))
13	Стандартная ошибка общей суммы	=(В4*В7*В12)/КОРЕНЬ(В6)
14	Степени свободы	=В6-1
15	<i>t</i> -значение	=СТЬЮДРАСПОБР(1-В8;В14)
16	Половина доверительного интервала	=В15*В13
17		

Окончание табл. ЕН.7.6

	А	В
18	Доверительный интервал	
19	Нижняя доверительная граница	=B11-B16
20	Верхняя доверительная граница	=B11+B16

При реализации этого шаблона ячейку В8 следует отформатировать так, чтобы величина 0,95 была представлена как 95% (см. врезку ЕР.5). Если при решении аналогичной задачи выборочное среднее и выборочное стандартное отклонение не известны и подлежат вычислению, необходимо заменить формулы в ячейках В5 и В7 формулами =СРЗНАЧ (*диапазон*) и =СТАНДОТКЛОН (*диапазон*) соответственно.

ЕН.7.7. Вычисление доверительного интервала, содержащего полную разность генеральной совокупности

Создадим два рабочих листа, предназначенных для вычисления доверительного интервала, содержащего общую сумму разностей элементов генеральной совокупности. В табл. ЕН.7.7–ЕН.7.9 приведены шаблоны двух рабочих листов, предназначенных для оценки общей суммы разностей между правильными и неправильными суммами накладных, как показано на рис. 7.14.

Таблица ЕН.7.7. Изменения рабочего листа Разности

	А	В
1	Разности	(D-Dbar)^2
2	9,03	=(A2-Оценка!\$B\$10)^2
3	7,47	=(A3-Оценка!\$B\$10)^2
4	17,32	=(A4-Оценка!\$B\$10)^2
...
11	7,43	=(A11-Оценка!\$B\$10)^2
12	2,99	=(A12-Оценка!\$B\$10)^2
13	4,63	=(A13-Оценка!\$B\$10)^2

В табл. ЕН.7.7 приведены формулы, которые необходимо добавить в рабочий лист Разности, находящийся в книге Chapter 7.XLS. Эти формулы вычисляют квадраты разностей между ошибками и средней выборочной разностью, которая отдельно вычисляется в ячейке В10 на втором рабочем листе. Под ошибкой подразумевается разность между правильной и неправильной суммами накладной. В свою очередь, формулы, записанные в ячейках Е11 и Е13 на рабочем листе Оценка (см. табл. ЕН.7.8), вычисляют количество ненулевых разностей и суммы их квадратов соответственно. Остальные ячейки в табл. ЕН.7.8 предназначены для вычисления стандартного отклонения разностей, записанного в ячейке В12 на листе Оценка (см. табл. ЕН.7.9).

Таблица ЕН.7.8. Шаблон рабочего листа Оценка (диапазон D9 : E16)

	С	D
9	Область вычислений	
10	Для стандартного отклонения разностей	
11	Количество ненулевых разностей	=СЧЕТ(Разности!A:A)
12	Количество нулевых разностей	=B5-E11
13	Сумма квадратов ненулевых разностей	=СУММ(Разности!B:B)
14	Сумма квадратов нулевых разностей	=E12*(-B10)^2
15	Сумма квадратов	=E13+E14
16	Дисперсия разностей	=E15/B15

Таблица ЕН.7.9. Шаблон рабочего листа Оценка (столбцы A и B, столбец C пуст)

	A	B
1	Вычисление доверительного интервала, содержащего общую сумму накладных	
2		
3	Данные	
4	Объем генеральной совокупности	5000
5	Выборочное среднее	100
6	Доверительный уровень	0,95
7		
8	Промежуточные вычисления	
9	Сумма разностей	=СУММ(Разности!A:A)
10	Средняя выборочная разность	=B9/B5
11	Полная разность	=B4*B10
12	Стандартное отклонение разностей	=КОРЕНЬ(E16)
13	Множитель	=КОРЕНЬ((B4-B5)/(B4-1))
14	Стандартная ошибка полной разности	=(B4*B12*B13)/КОРЕНЬ(B5)
15	Степени свободы	=B5-1
16	t-значение	=СТЮДРАСПОБР(1-B6;B15)
17	Половина доверительного интервала	=B16*B14
18		
19	Доверительный интервал	
20	Нижняя доверительная граница	=B11-B17
21	Верхняя доверительная граница	=B11+B17

Второй шаблон предназначен для вычисления половины длины доверительного интервала. Результат вычислений записывается в ячейку В17 и представляет собой произведение t -значения, возвращаемого функцией СТЬЮДРАСПОБР (1-доверительный_уровень; степени_свободы) и стандартной ошибки полной разности. Добавляя и вычитая полученную величину из полной разности, мы получаем верхнюю и нижнюю границы доверительного интервала.

Реализуя шаблон, представленный в табл. ЕН.7.7, введите формулу в ячейку В2 и скопируйте ее во всей ячейки этого столбца вплоть до строки 13. При реализации шаблона, представленного в табл. ЕН.7.9, отформатируйте ячейку В6 так, чтобы величина 0,95 была представлена как 95% (см. врезку ЕР.5). Для решения аналогичных задач измените первый шаблон, введя в столбце А новые разности. Если в новой задаче больше 12 разностей, скопируйте формулу из ячейки В13 в ячейки, лежащие ниже. Если в новой задаче меньше 12 разностей, удалите лишние строки.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Cochran, W. G., *Sampling Techniques*, 3rd ed. (New York: Wiley, 1977).
2. Fisher, R. A., and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 5th ed. (Edinburgh, Scotland: Oliver & Boyd, 1957).
3. Kirk, R. E., ed., *Statistical Issues: A Reader for the Behavioral Sciences* (Belmont, CA: Wadsworth, 1972).
4. Lasen, R. L., and M. L. Marx, *An Introduction to Mathematical Statistics and Its Applications*, 2nd ed. (Englewood Cliffs, NJ: Prentice Hill, 1986).
5. *Microsoft Excel 2002* (Redmond, WA: Microsoft Corporation, 2001).
6. Snedekor, G. W., and W. G. Cochran, *Statistical Methods*, 7th ed. (Ames, IA: Iowa State University Press, 1980).